ED 454 279   TM 032 888

AUTHOR   Parshall, Cynthia G.; Kromrey, Jeffrey D.; Harmes, J.
                  Christine; Sentovich, Christina
TITLE   Nearest Neighbors, Simple Strata, and Probabilistic
                  Parameters: An Empirical Comparison of Methods for Item
                  Exposure Control in CATs.
SPONS AGENCY   Defense Manpower Data Center, Monterey, CA.
PUB DATE   2001-04-00
NOTE   60p.; Paper presented at the Annual Meeting of the National
                  Council on Measurement in Education (Seattle, WA, April
                  11-13, 2001).
CONTRACT   M67004-00-D-0002
PUB TYPE   Numerical/Quantitative Data (110) -- Reports - Research
                  (143) -- Speeches/Meeting Papers (150)
EDRS PRICE   MF01/PC03 Plus Postage.
DESCRIPTORS   Ability; *Adaptive Testing; *Computer Assisted Testing;
                  Estimation (Mathematics); Item Banks; Monte Carlo Methods;
                  *Test Items
IDENTIFIERS   *Item Exposure (Tests)

ABSTRACT
          Computerized adaptive tests (CATs) are efficient because of
their optimal item selection procedures that target maximally informative
items at each estimated ability level. However, operational administration of
these optimal CATs results in a relatively small subset of items given to
examinees too often, while another portion of the item pool is almost unused.
This situation both wastes a portion of the available items and can be a
security risk. A number of exposure control methods have been developed to
reduce this effect. In this study, the effectiveness of three methods was
investigated in comparison to baseline conditions of No Control and Random
item selection. These procedures were: (1) the Sympson-Hetter method (J.
Sympson and R. Hetter, 1985); (2) the Nearest Neighbor method (R. Holmes and
D. Segall, 1999); and (3) Stratified-a methods (H. Chang and Z. Ying, 1997).
Using Monte Carlo procedures, these methods were examined under varying
target maximum exposure rates. Results are reported in terms of pool usage,
test precision and bias, both unconditionally and conditionally. Three
methods were completely successful in preventing marginal administration
rates beyond the specified target maximum, the Sympson Hetter and Nearest
Neighbor methods and the Stratified-a method incorporating item freezing.
(Contains 26 figures and 18 references.) (Author/SLD)

Nearest Neighbors, Simple Strata, and Probabilistic Parameters:
An Empirical Comparison of Methods for Item Exposure Control in CATs

Cynthia G. Parshall
Jeffrey D. Kromrey
J. Christine Harmes
Christina Sentovich
*Institute for Instructional Research &Practice, and
Educational Measurement & Research Department,
University of South Florida*

Abstract

Computerized adaptive tests are efficient because of their optimal item selection procedures that target maximally informative items at each estimated ability level. However, operational administration of these optimal CATs results in a relatively small subset of items given to examinees overly often, while another portion of the item pool is almost unused. This situation both wastes a portion of the available items and can be a security risk. A number of exposure control methods have been developed to reduce this effect. In this study, we investigate the effectiveness the Sympson-Hetter, Nearest Neighbor, and Stratified-a methods in comparison to baseline conditions of No Control and Random item selection. Using Monte Carlo procedures, we examine these methods under varying target maximum exposure rates. Results are reported in terms of pool usage, test precision and bias, both unconditionally and conditionally.

Nearest Neighbors, Simple Strata, and Probabilistic Parameters:
An Empirical Comparison of Methods for Item Exposure Control in CATs

When items are selected during a computerized adaptive test (CAT) based solely on their measurement properties, item pool usage is found to be very uneven. Operational administrations have found that a relatively small subset of items is administered with an undesired high frequency, while another portion of the item pool is almost unused. This both wastes a portion of the available items and, even more importantly, it clearly presents a security risk for testing programs that are available on various occasions throughout the year. A number of exposure control methods have been developed to reduce this effect.

The Sympson-Hetter method (Sympson & Hetter, 1985) was one of the earliest approaches to control item over-exposure, and a number of adaptations of this method have been developed (e.g., Davey & Parshall, 1995; Nering, Davey & Thompson, 1998; Holmes & Segall, 1999; Parshall, Davey, & Nering, 1998; Parshall, Kromrey, & Hogarty, 2000; Stocking & Lewis, 1995; Thomasson, 1995). In all of these probabilistic approaches to exposure control, a series of simulations is conducted to assign a unique *exposure parameter* to each item. This parameter is used to probabilistically limit the frequency with which a selected item is administered. These methods have been found to be reasonably effective, but they can be cumbersome to implement. Furthermore, every time a change is made to the item pool, the preparatory simulations must be conducted again.

The Nearest-Neighbor method (Holmes & Segall, 1999), is an extension of the Sympson-Hetter approach that attempts to equalize exposure rates across items that are similar in level of information and performance. Based on the Sympson-Hetter exposure control parameters, item usage rates are simulated and items are sorted by their usage. Items are then grouped by calculating a distance parameter based on item information functions, and establishing the "nearest neighbors", beginning with the most used item. A smoothing algorithm is applied to adjust exposure rates within each group. This procedure is carried out until all items have been smoothed or a specified stopping rule for item usage has been reached. This method was shown to be successful in increasing the number of item combinations that would be presented to examinees with minimal reduction to test information. However, it retains the weaknesses of other probabilistic methods of complexity and being subject to item pool changes.

A very different approach is taken in the Stratified-a method (Chang & Ying, 1997). No simulations or exposure parameters are used; rather, the items in a pool are assigned to strata, based on their a-values, an estimate of the item's discriminatory power. Early in the test, items are administered from the stratum with the lowest a-parameters. As the test progresses, the strata with higher a-values are used. Extreme overuse of some items can still be found under this method; however, only a small number of items tend

to be overused (Parshall, Kromrey, & Hogarty, 2000). An adaptation of the Stratified-a method that appears to address this problem is to temporarily render items unavailable for selection when they exceed a target administration rate – that is, to "freeze" these items in the selection algorithm until their administration rate drops below the target value (Kromrey, Parshall & Harmes, 2000; Parshall, Kromrey, & Harmes, 2000).

Purpose

Although theoretically sound, the Sympson-Hetter is computationally complicated and logistically involved. Further, it may provide an inadequate degree of exposure control for many applications. The Nearest Neighbor method builds on the Sympson-Hetter method, adding to its effectiveness but also to its complexity. The Stratified-a method, in contrast, is straightforward and easy to implement, but may provide exposure control to a lesser extent than the more complex methods. The variation of the Stratified-a method that temporarily freezes items might address this weakness, while retaining the advantages of the method. The purpose of the study was to empirically investigate the Sympson-Hetter and Nearest Neighbor methods along with controlled experimental variations of item freezing in conjunction with the Stratified-a method.

Methods

For this research the Sympson-Hetter, Nearest Neighbor, and Stratified-a (with and without freezing) exposure control methods were all implemented in a Monte Carlo study in which adaptive testing was simulated under controlled conditions. The effectiveness of the two variations of the Stratified-a exposure control method and the Sympson-Hetter and Nearest Neighbor method were compared to each other and to two additional "baseline" conditions (No Control and completely Random item selection). These six exposure control methods were investigated at four target maximum exposure rates (.15, .25, .33, and .40), resulting in 24 study conditions.

Exposure Control Methods Operationalized

Specific implementation decisions and steps are needed for most exposure control methods. For the probabilistic methods, a preliminary simulation phase is necessary. For the Sympson-Hetter method, the exposure control parameters were initialized to a value close to the target maximum exposure rate. These values were then free to either increment or decrement, depending upon the frequency with which their associated items were administered. A series of 600 simulation cycles of 5,000 exams each was conducted. The final Sympson-Hetter exposure control parameters resulting from this process were saved, for use in the "operational testing" phase of the simulation.

Preparation of the Nearest Neighbor exposure parameters involved further adjusting the Sympson-Hetter parameters in a "smoothing" process. Following the procedures suggested by Holmes and Segall (1999), items were sorted by administration rate and item "neighbors" (i.e., those having distances of .20

or less) were clustered. The exposure parameters within each cluster (or neighborhood) were then smoothed. A set of 6 iterations was conducted, in which 5000 exams were administered, followed by further smoothing.

For the Stratified-a method, no preliminary simulation was needed, but several decisions relative to the item pool were made. In this case, the item pool was divided into four strata, with four items to be drawn from each of the first three strata, and three items from the final strata (resulting in a test length of 15-item). Within the specified strata an item is usually selected based on how close its b-value is to the examinee's current estimate of theta. However, the first four items (i.e., the entire first strata) of each test were selected randomly from within the initial stratum. Since the simulated CAT began each test assuming an examinee's ability was 0, this modification was incorporated into the Stratified-a method to avoid all examinees being presented with near identical items early in the test.

In the "freeze" condition of the Stratifed-a method items that exceeded a target administration rate were "frozen", or rendered temporarily unavailable for selection. As more tests were administered, this proportional administration rate for the frozen items dropped below the target rate again; at this point the frozen items were "thawed", and once again were available for selection and use. In the "no freeze" condition, items were selected and administered using the Stratified-a method without the augmentation of temporary freezing.

For the Random method and the No Control method, no preparations of these sorts were necessary. For both the Sympson-Hetter and the Nearest Neighbor methods, the study condition of "target exposure rate" was manipulated in this preparatory phase of the study; for all of the methods, its effect was investigated in the next phase. In this next "operational simulation" phase, adaptive test administrations were simulated for 50,000 examinees in each study condition.

CAT Characteristics

The CAT characteristics defined for this study were intended to reflect administration of the Arithmetic Reasoning (AR) test of the computerized adaptive Armed Services Vocational Aptitude Battery (CAT-ASVAB). An item pool consisting of 187 AR items was used to generate fixed-length CATs of 15-items each. No content constraints were imposed on the item selection procedures. Provisional ability estimates were computed by Owen's Bayes mode approximation (Owen, 1969, 1975), while final estimates were obtained using MAP.

Item selection was managed differently depending upon the study condition. The Random method had no limitations on item selection; rather, each item was drawn randomly from the pool. The No Control method used maximum information (MI) item selection, with no exposure control. The Sympson-Hetter and Nearest Neighbor methods also used MI, incorporating their own exposure control parameters as limiting factors.

Data Generation

Simulated item responses were generated based on operational item parameter estimates and a multidimensional item response theory (MIRT) model. This model included not only the major dimensions that provide basic structure, but also numerous minor dimensions that are characteristic of actual data. MIRT data generation provides simulated data that are more similar to real data than those produced by more typical unidimensional IRT models (Davey, Nering, & Thompson, 1997; Parshall, Kromrey, Chason, & Yi, 1997).

Existing 3-PL item parameter estimates for the set of operational ASVAB AR items were used to generate 20,000 examinee responses to all 187 items. These simulated data were then analyzed using the program Noharm (Fraser & McDonald, 1986) to obtain item parameters calibrated in a 6-dimensional space. The set of MIRT item parameters were used along with simulated examinee abilities to generate examinee responses to the adaptive tests. Item responses were generated by determining the probability of a correct response on a given item, for a given examinee, and then comparing that probability to a random number sampled from a uniform (0,1) distribution. If the probability of a correct response was greater than the random number then the response was scored correct; otherwise, the response was scored incorrect.

Effectiveness Criteria

The relative effectiveness of the exposure control methods was evaluated by examining multiple criteria. The success of the methods in controlling item exposure was investigated by computing the administration rates for items both marginally (for the overall sample of 50,000 simulated examinees) and conditional on examinee ability. The simulation conditions that applied no exposure control and random item selection provided reference points against which the administration rates under exposure control could be checked.

Further, the use of exposure control methods influences both the accuracy and precision of examinee ability estimates. Bias in each ability estimate was calculated as the simple difference between the estimated ability and true ability. Because true ability was defined in the space of the 6-dimensional MIRT model used for data generation, a unidimensional theta value that most closely approximates the MIRT ability vector was calculated for each simulated examinee. This served as the best unidimensional representation of true ability. The method suggested by Fan, Thompson, and Davey (1999) was used in this step. In this approach, the unidimensional theta value that minimizes the sum of squared differences in item response probabilities, across the entire item pool, between the unidimensional theta and the 6-dimensional MIRT vector of true theta values was computed. Finally, precision in ability estimates was evaluated by computing the posterior variance of each ability estimate in the simulations.

Results

The results are reported in terms of pool-usage, and ability estimation error and bias. One goal of this line of research has been to develop good methods of examining item exposure performance. A variety of figures are used to help satisfy this goal.

Pool Usage

Pool usage information is displayed in several figures. The entire distribution of marginal item administration rates is shown in Figures 1a-d for the target maximum exposure rates of .15, .25, .33, and .40 respectively. The pattern of results for the six exposure control conditions are similar across the target maximum exposure rates. Note that the Random method shows ideal pool usage, without problems of either over-exposure or under-exposure, while the No Control condition shows problems with both. The results also clearly show that the inclusion of freezing in the Stratified-a method is both necessary and effective in dealing with over-exposure, and also appears to help address under-exposure. Finally, the Sympson-Hetter and Nearest Neighbor display very similar administration rate distributions.

Another visual examination of pool usage is considered next. The target maximum exposure rates can be regarded as test security criteria for item administration rates that a testing program might establish as a goal. While these rates are used directly in the preliminary simulation phase of probabilistic methods such as the Sympson-Hetter and Nearest Neighbor, they may be used as indirect goals for any method. If an exposure control method allows an item to be administered more frequently than this target, the item may be considered to have been over-exposed. A complementary goal in the use of exposure control is to improve pool usage; thus, items may also potentially be under-exposed. For this study, an item is classified as under-exposed if it is administered less than half the times it would be given under completely random item administration. For a test length of 15 and a pool size of 187, an item with no restrictions might be administered roughly 8% of the time; half of that completely random administration would be approximately 4%. Thus, any item used on 4% of the exams or fewer is counted as underexposed. While the criteria for under-exposure is consistent for a given test length and pool size, the criteria for over-exposure is dependent upon the target maximum exposure rate; in this study, four target rates were investigated.

The proportion of items over- and under-exposed is displayed in Figures 2 a-d, for each exposure control method across target exposure rate. Note that No Control shows the worst performance, with a few items over-exposed and many items under-exposed across all four target rates, and Random shows the best performance, with no instances of either under- or over-exposure. For the remaining methods (which are more appropriate for actual operational use), it can be noted that under-exposure, or poor pool usage, is more of an issue with relaxed target rates (e.g., .40) than with stringent ones (e.g., .15). This is an expected trend, given that the use of stringent exposure control severely limits the availability of those

items in the pool that are highly desirable to the item selection algorithm. This forces other items to be used and thus improves overall pool usage.

The only method in which over-exposure remains a problem is the standard Stratified-a – which has no inherent direct control over item administration rates. The inclusion of freezing to the Stratified-a removes any over-exposure problem and concomitantly reduces the under-exposure problem. The Stratified-a-with-freezing, the Sympson-Hetter, and the Nearest Neighbor perform very similarly to one another, displaying no problem with over-exposure, and only a moderate problem with under-exposure.

A conditional view of pool usage in displayed in Figures 3 a-d. This information shows the 95th percentile of the distribution of item administration rates, conditional on ability. In other words, at each level of ability a value close to the "maximum" item administration rate is plotted; 95% of the items were administered at that ability level less often than the plotted point. These figures differ from the earlier ones in that the relative performance of the exposure control methods across ability levels is shown. The Random method has the lowest item administration rates across ability, as would be expected. On the other extreme, the No Control method shows the highest item administration rates across ability. The remaining, more realistic, methods fall between these two. The Sympson-Hetter and Nearest Neighbor methods perform almost identically to one another, maintaining conditional item administration rates close to each target rate, across most of the ability range. For most of the ability range, and all target maximum rates, the poorest performance is again displayed by the standard Stratified-a method due to that method's lack of direct control of item administration rates. Considerably better performance can be seen by the Stratified-a-with-freezing method. This adapted Stratified-a method performs very similarly to the Sympson-Hetter and Nearest Neighbor methods at more stringent target rates, but shows somewhat higher conditional administration rates with more relaxed target rates.

Test Precision

Test precision is investigated in this study by an examination of the error variance of the final ability estimates. These posterior variances of the ability estimates, conditional on true ability, are provided for all of the study conditions in Figures 4 a-d. Similar patterns of results are displayed across the four target rates. All of the methods display greater error variance in the tails of the ability distribution, where less information is available in the item pool. While the methods perform fairly similarly across most of the range of ability estimates, distinct differences are notable, particularly near the center of the ability range. In that area, the smallest marginal error variance is found, as expected, for the No Control condition, and the largest marginal error variance is found for the Random method. The Sympson-Hetter and Nearest Neighbor methods again perform very similarly across ability. Additionally, the Stratified-a and Stratified-a-with-freezing methods perform similarly to one another, indicating that the inclusion of freezing did not lessen the accuracy of the ability estimation.

Test Bias

Bias in the ability estimates, computed as the simple difference between the estimated ability and true ability, is plotted in Figures 5 a-d. The overall pattern of results displayed, in which positive bias is seen at low ability estimates and negative bias is seen at high ability estimates, is typical of Bayesian ability estimation methods. As a whole, the methods perform similarly to one another, and similarly across the four target maximum exposure rates.

Freeze Rates

Finally, plots of the frequency with which each item is frozen are provided in Figures 6 a-d, by a-parameter and b-parameter, for the four target maximum conditions. Every item in the pool is plotted as a circle in these figures; the more frequently an item was frozen, the larger the size of that item's circle. It is evident that items with b-values in the vicinity of 0, and with a-values over 1.0, tended to be frozen more frequently. These middle-difficulty, high-discrimination items were apparently in great demand, resulting in their tendency to be frozen at higher rates.

Summary

Any CAT program must be a compromise between competing goals. They can be efficient, allowing for the selection of items that provides optimal measurement at each examinee's estimated level of ability, thereby maximizing efficiency and accuracy. However, this efficiency results in very uneven item pool usage. In addition to the economic concern of items that are used too rarely, frequently administered items can become compromised, at which point they no longer provide valid measurement. The need for exposure control is clear.

For the 187-item pool investigated in this research, three of the exposure control methods were completely successful in preventing marginal administration rates beyond the specified target maximum, even with a target as low as .15: the Sympson-Hetter, Nearest Neighbor, and Stratified-a method incorporating item freezing. When the Stratified-a method was implemented without freezing, a small number of items were administered at excessively high rates. The impact of freezing was especially evident in the examination of administration rates conditional on examinee ability. These results are consistent with those of earlier studies that incorporated a larger pool and a longer fixed length adaptive test (Kromrey, Parshall & Harmes, 2000; Parshall, Kromrey, & Harmes, 2000). Such findings suggest that the Stratified-a method with freezing, incorporating a simple non-probabilistic exposure control strategy, appears to do remarkably well at constraining item administration rates to their target maximum goals, without degrading test precision unacceptably.

In this study, the Nearest Neighbor method performed very similarly to the Sympson-Hetter. While close performance is to be expected, given that the Nearest Neighbor exposure parameters are smoothed values of the Sympson-Hetter parameters, more distinction might have been seen if the number of

smoothing iterations were increased. However, with the test conditions simulated in this study, both methods were effective in controlling item administration rates.

One limitation of this study, as in many CAT simulations studies, is that methods are investigated within specific test definitions. In this case, a short, fixed-length test, without content constraints, was administered from a very informative pool. This may have lessened the extent to which exposure control methods had an impact on test precision or bias. The use of exposure control methods in test context with smaller, less informative pools is likely to present greater challenges for item exposure control and successful exposure control is likely to evidence a cost in terms of bias and larger standard errors.

References

Chang, H. & Ying, Z. (1997). *A-stratified multistage computerized adaptive testing*. Applied Psychological Measurement.

Davey, T., Nering, M. L., & Thompson, T. (1997, June). *Realistic simulation of item response data*. Paper presented at the annual meeting of the Psychometric Society, Gatlinburg, TN.

Davey, T., & Parshall, C. G. (1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Fan, M., Thompson, T., & Davey, T. (1999, April). *Constructing adaptive tests to parallel conventional programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Fraser, C., & McDonald, R. P. (1986). NOHARM II: A FORTRAN program for fitting unidimensional and multidimensional normal ogive models of latent trait theory [computer program]. Armidale, New South Wales, Australia: Center for Behavioral Studies, The University of New England.

Holmes, R. M., & Segall, D. O. (1999, April). *Reducing item exposure without reducing precision (much) in computerized adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal.

Kromrey, J. D., Parshall, C. G. & Harmes, J. C. (2000, November). *Variations on stratification: The use of item freezing to augment the Stratified-a exposure control method*. Paper presented at the annual meeting of the Florida Educational Research Association, Tallahassee.

Nering, M., Davey, T., & Thompson, T. (1997, June). *Simulation of realistic ability vectors*. Paper presented at the annual meeting of the Psychometric Society, Gatlinburg, TN.

Owen, R. J. (1969). *A Bayesian approach to tailored testing* (Research Report 69-92). Princeton, NJ: Educational Testing Service.

Owen, R.J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association, 70*, 351-356.

Parshall, C. G., Davey, T., & Nering, M. L. (1998). *Test development exposure control for adaptive testing*. Paper presented at the symposium Adaptive Testing Research at ACT at the annual meeting of the National Council on Measurement in Education, San Diego.

Parshall, C. G., Kromrey, J. D., Chason, W. & Yi, Q. (1997) *Evaluation of parameter estimation under modified IRT models and small samples*. Paper presented at the annual meeting of the Psychometric Society, Gatlinburg, TN.

Parshall, C. G., Kromrey, J. D., & Harmes, J. C. (2000, July). *Variations on Stratification Methods for Item Exposure Control: Effects of Suspending Selection and Sharpening Strata*. Paper presented at the annual meeting of the Psychometric Society, Vancouver.

Parshall, C. G., Kromrey, J. D., & Hogarty, K. Y. (2000, April). *Sufficient simplicity or comprehensive complexity? A comparison of probabilistic and stratification methods of exposure control*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Reckase, M. D., Thompson, T., & Nering, M. . (1997, June). *Identifying similar item content clusters on multiple test forms.*. In T. Miller (chair) High-dimensional simulation of item response data for CAT research. Symposium conducted at the annual meeting of Psychometric Society, Gatlinburg, TN.
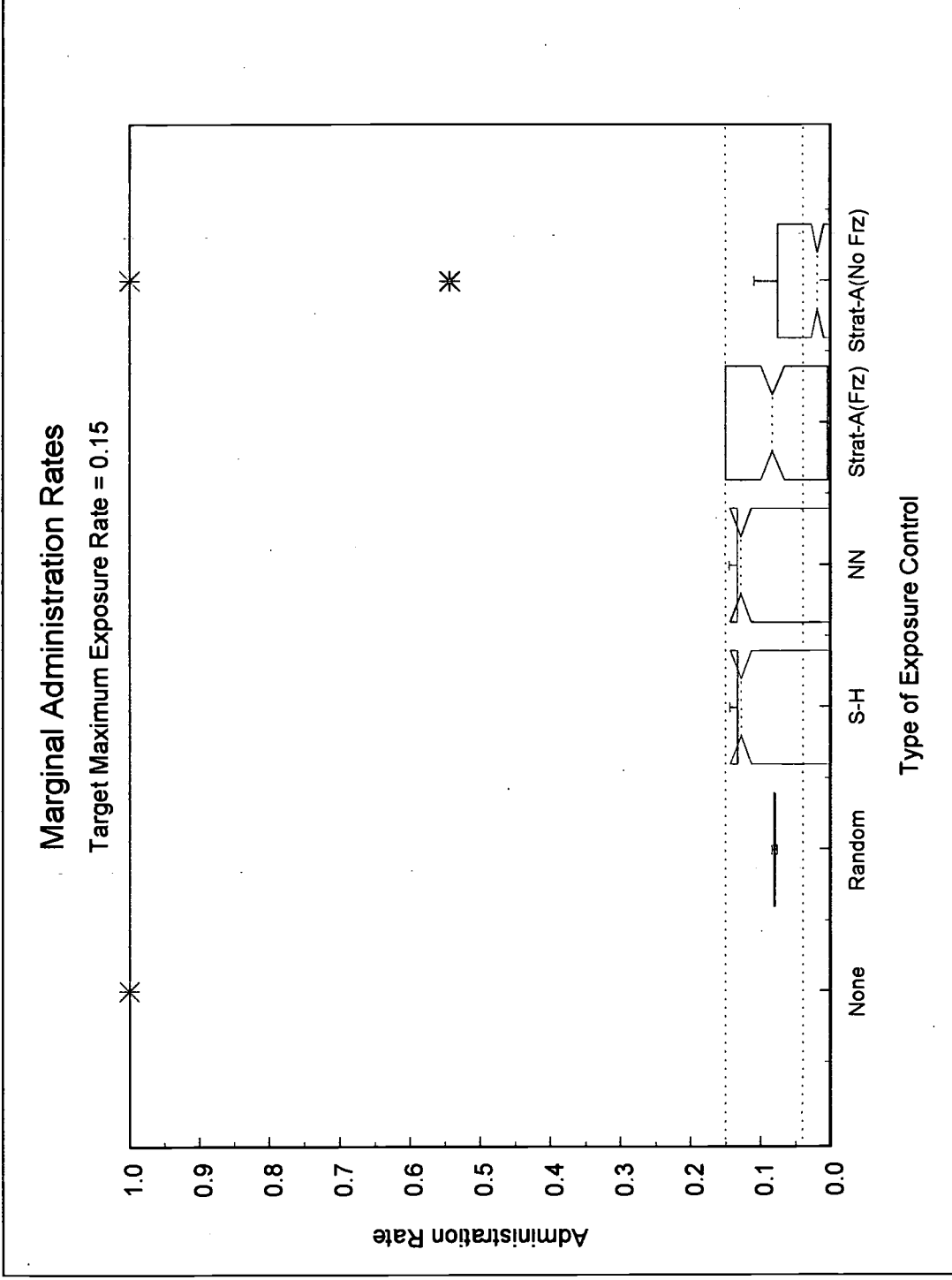
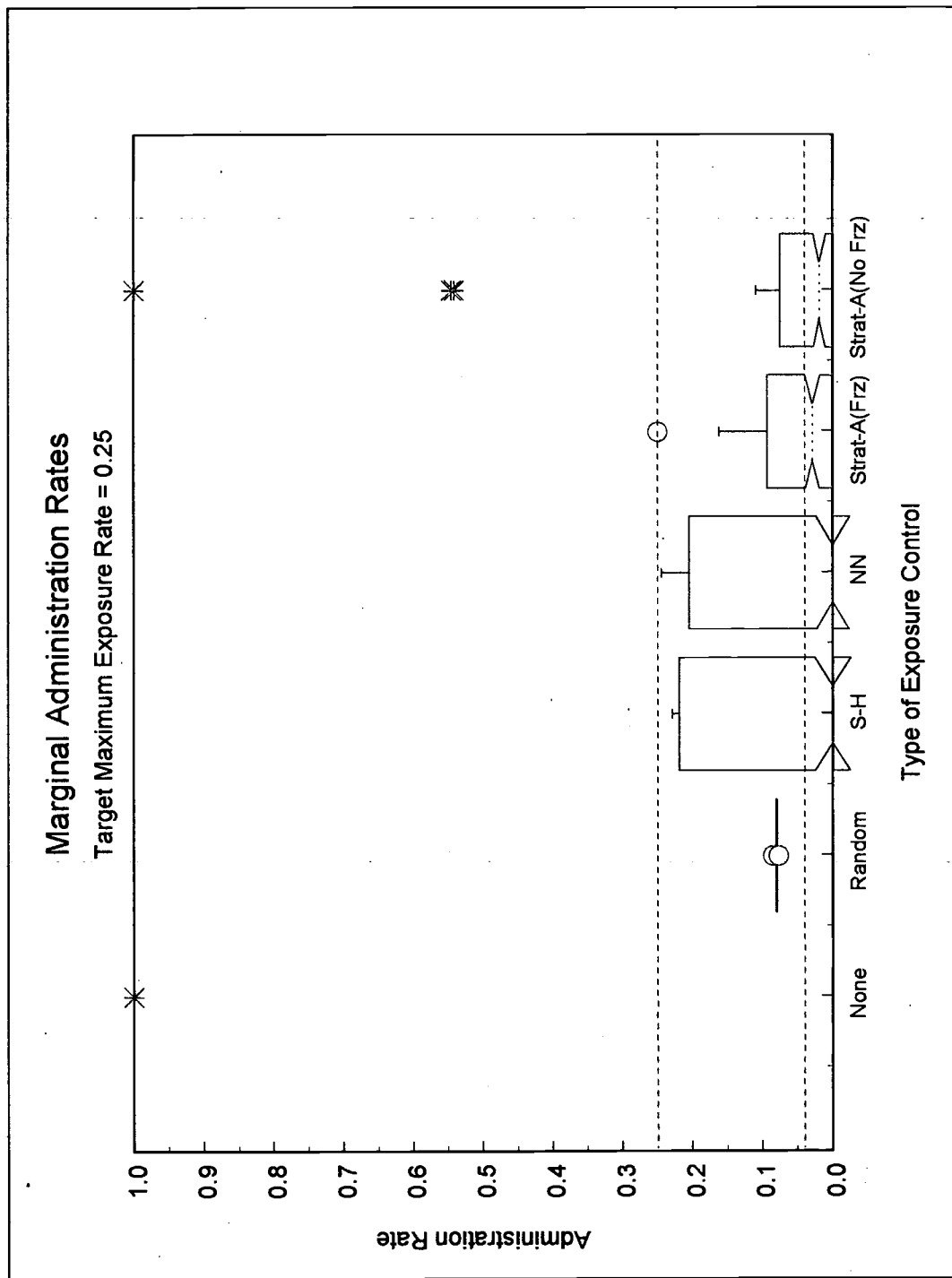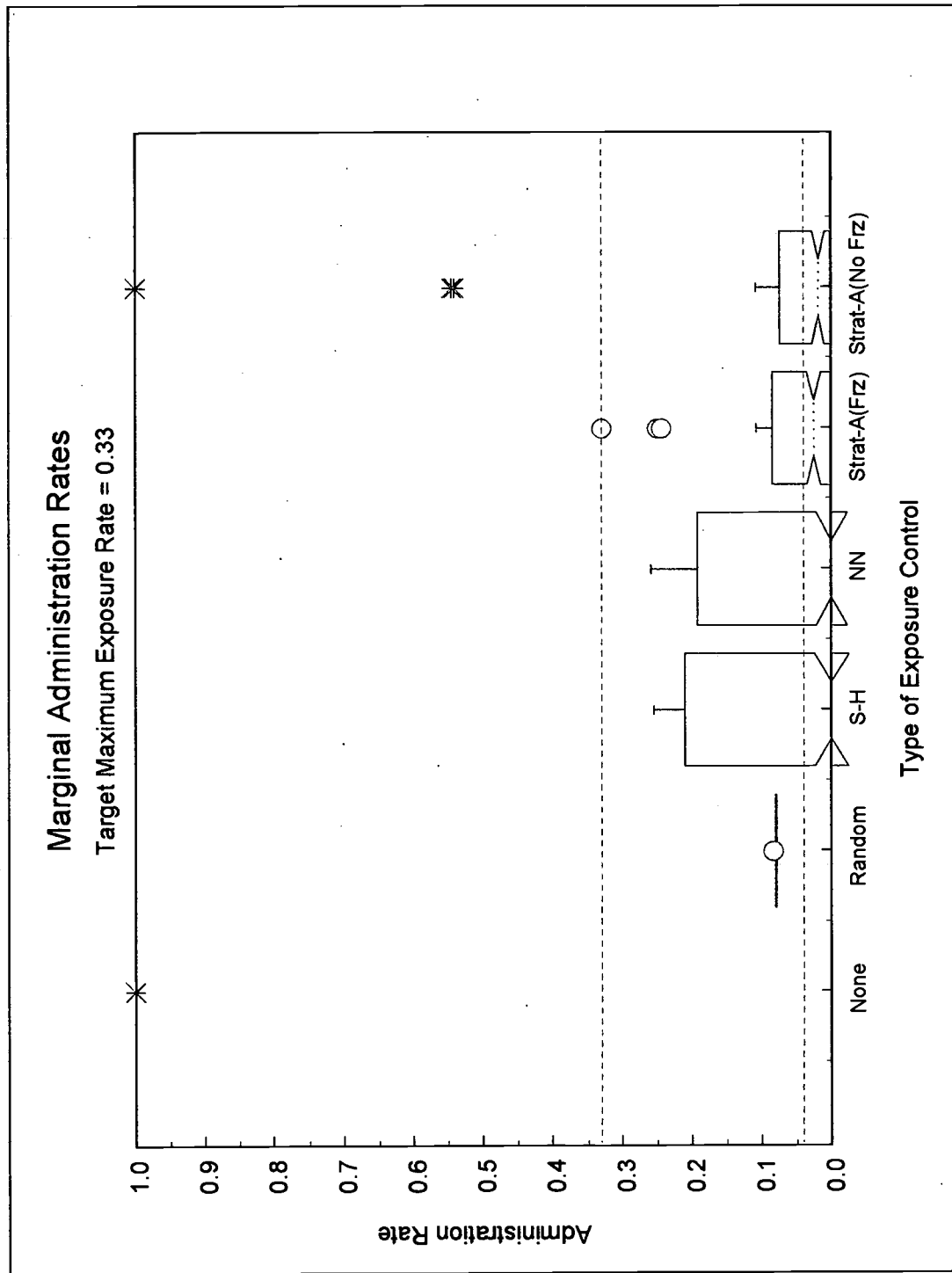Stocking, M. L., & Lewis, C. (1995). *Controlling item exposure conditional on ability in computerized adaptive testing.* (Research Report 95-24). Princeton, NJ: Educational Testing Service.

Sympson, J.B. & Hetter, R.D. (1985). Controlling item-exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association* (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.

Thomasson, G. L. (1995). *New item exposure control algorithms for computerized adaptive testing.* Paper presented at the annual meeting of the Psychometric Society, Minneapolis.

Thompson, T., Nering, M., & Davey, T. (1997, June). *Multidimensional IRT scale linking without common items or common examinees.* In T. Miller (chair) High-dimensional simulation of item response data for CAT research. Symposium conducted at the annual meeting of Psychometric Society, Gatlinburg, TN.

Marginal Administration Rates
Target Maximum Exposure Rate = 0.15

Figure 1 a

16

Marginal Administration Rates
Target Maximum Exposure Rate = 0.25

Figure 1 b

15

Marginal Administration Rates
Target Maximum Exposure Rate = 0.33

Figure 1 c

17   18

Figure 1 d

20
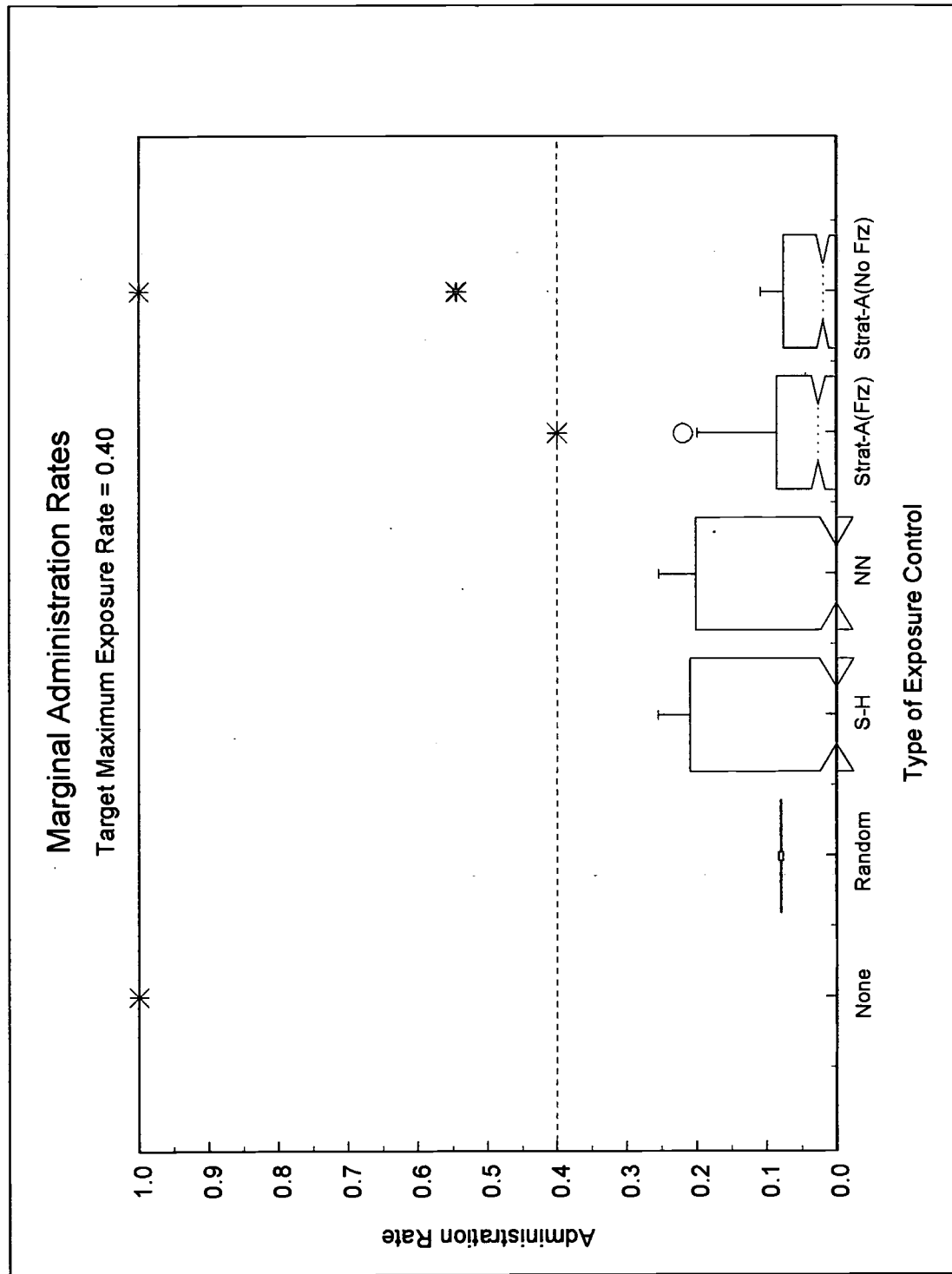
Proportion of Items Over and Under Exposed (Target Rate = 0.15)
By Exposure Control Method

Figure 2 a

21

22
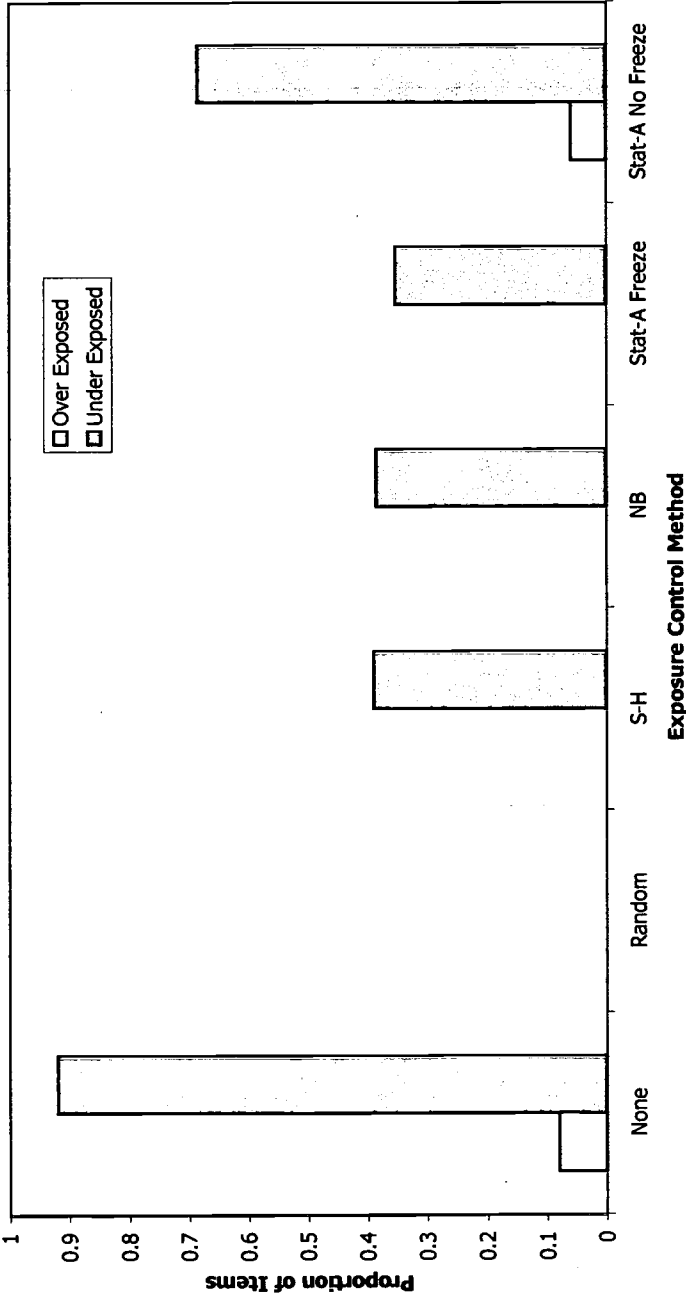
Proportion of Items Over and Under Exposed (Target Rate = 0.25)
By Exposure Control Method
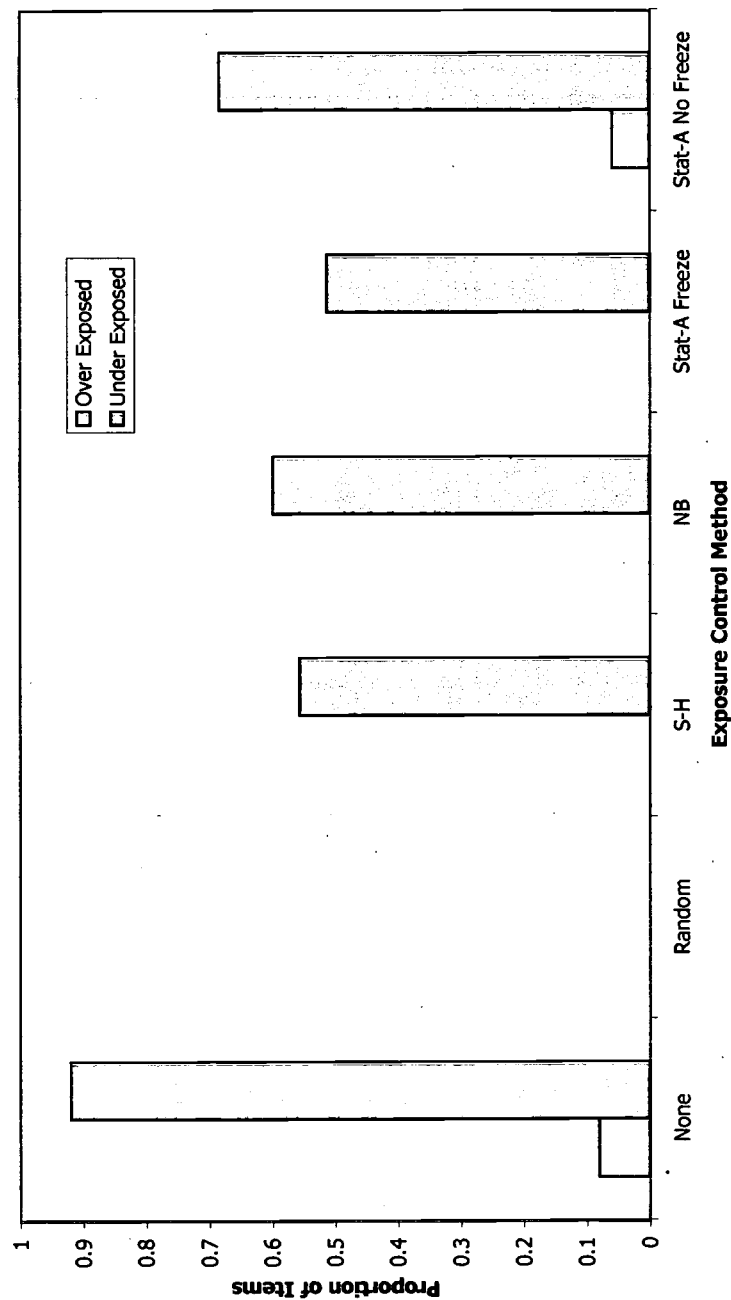
Figure 2 b

Proportion of Items Over and Under Exposed (Target Rate = 0.33)
By Exposure Control Method
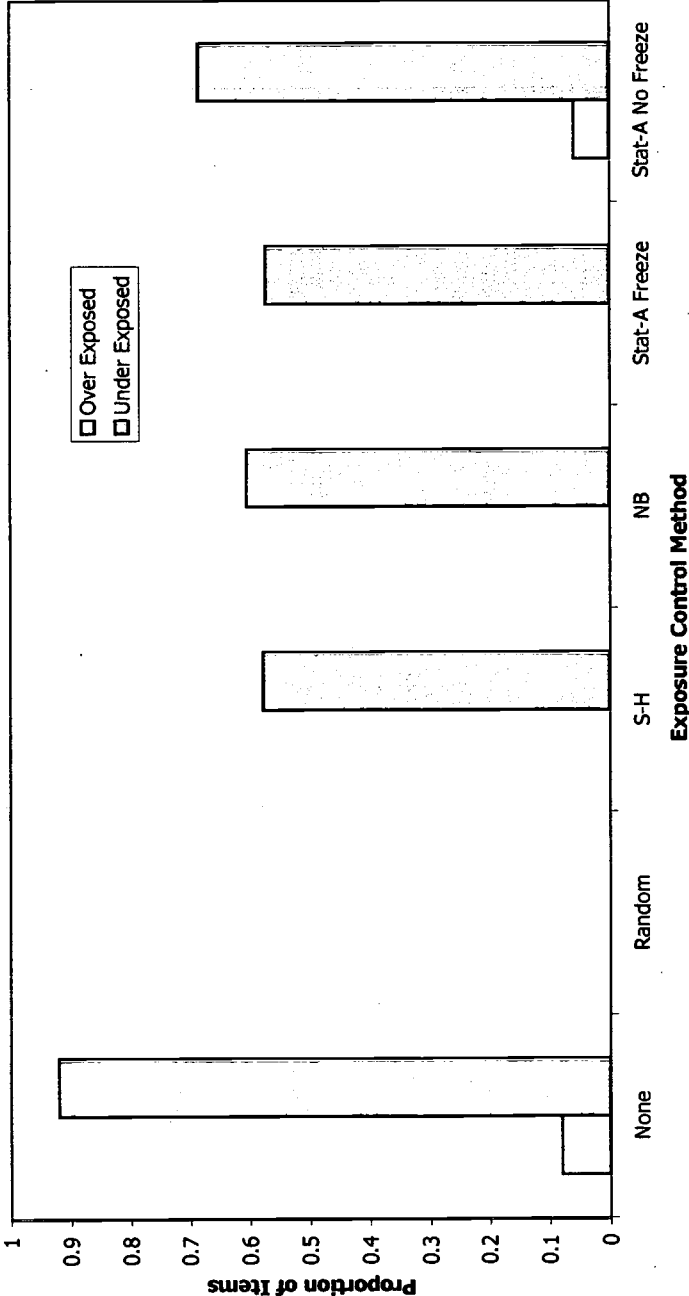
Figure 2 c
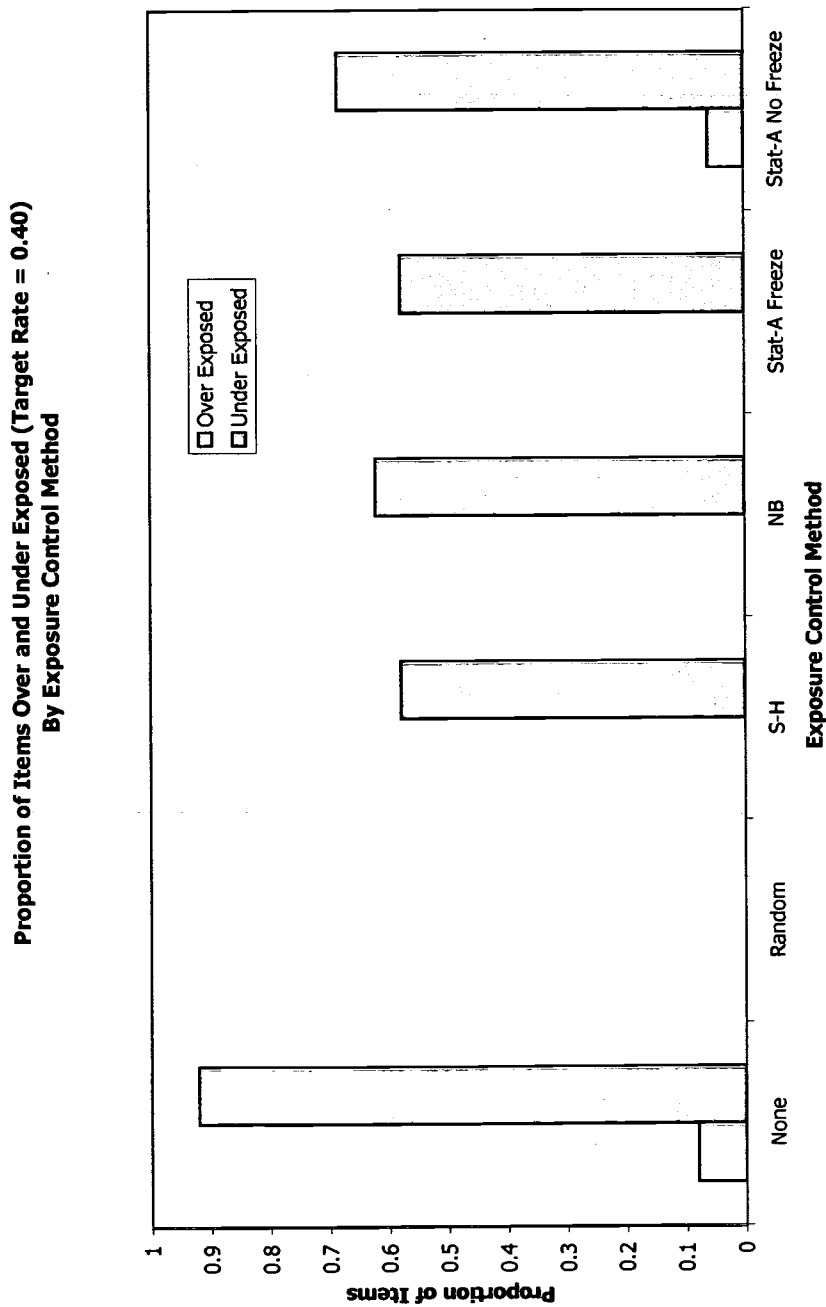
26

Proportion of Items Over and Under Exposed (Target Rate = 0.40)
By Exposure Control Method

Figure 2 d

28

95th Percentiles of Conditional Exposure Distributions Target Maximum Exposure Rate = 0.15

Figure 3 a

29

30
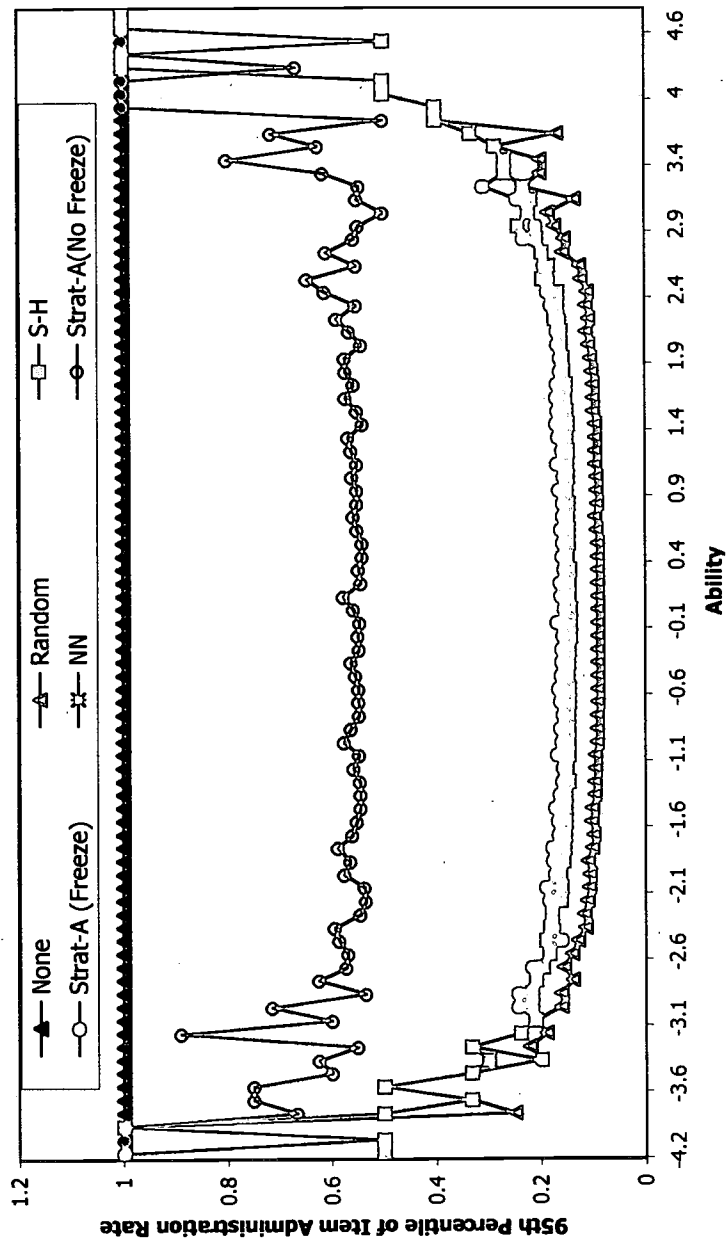
95th Percentiles of Conditional Exposure Distributions Target Maximum Exposure = 0.25
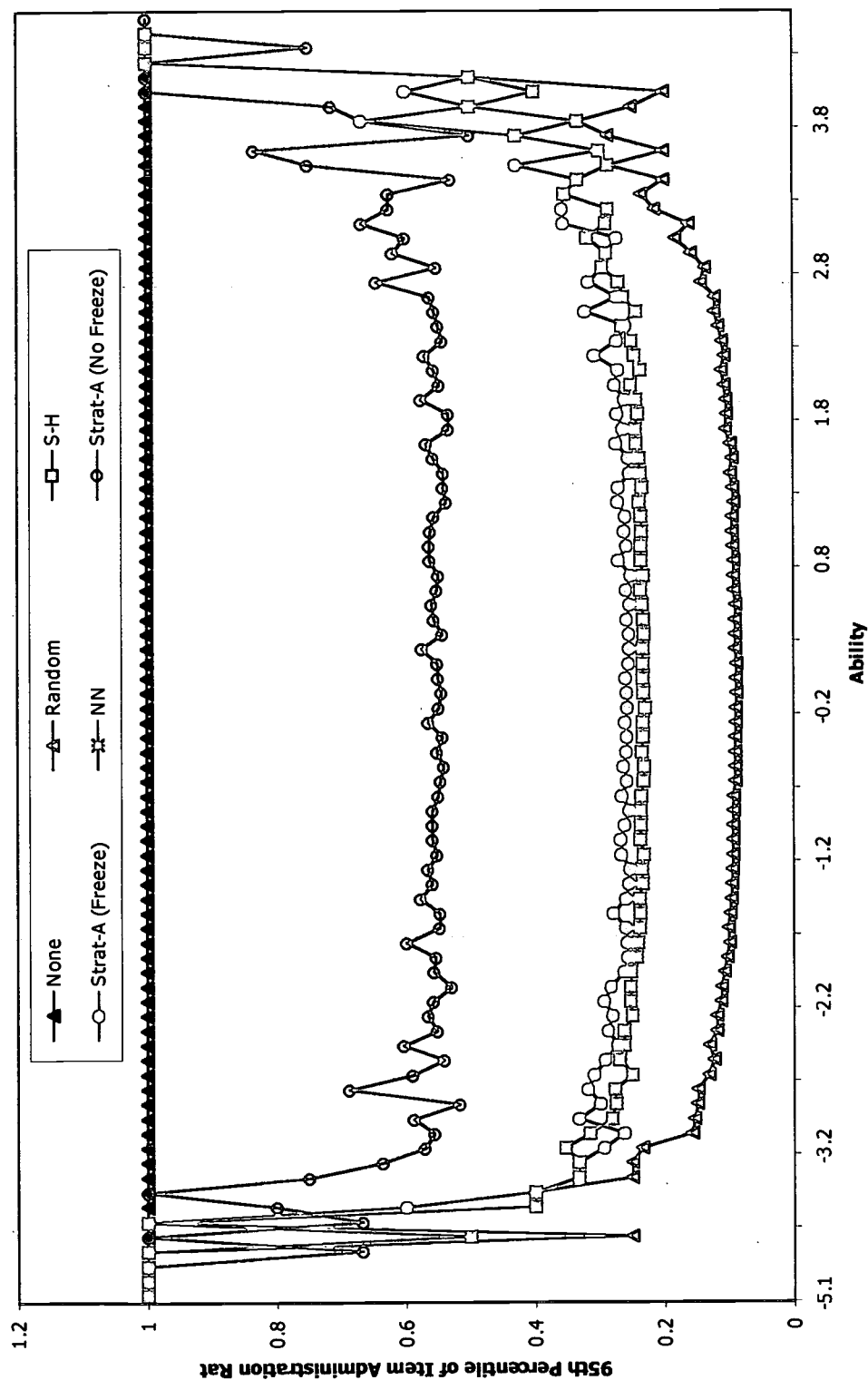
Figure 3 b

95th Percentiles of Conditional Exposure Distributions Target Maximum Exposure Rate = 0.33

Figure 3 c

33

34
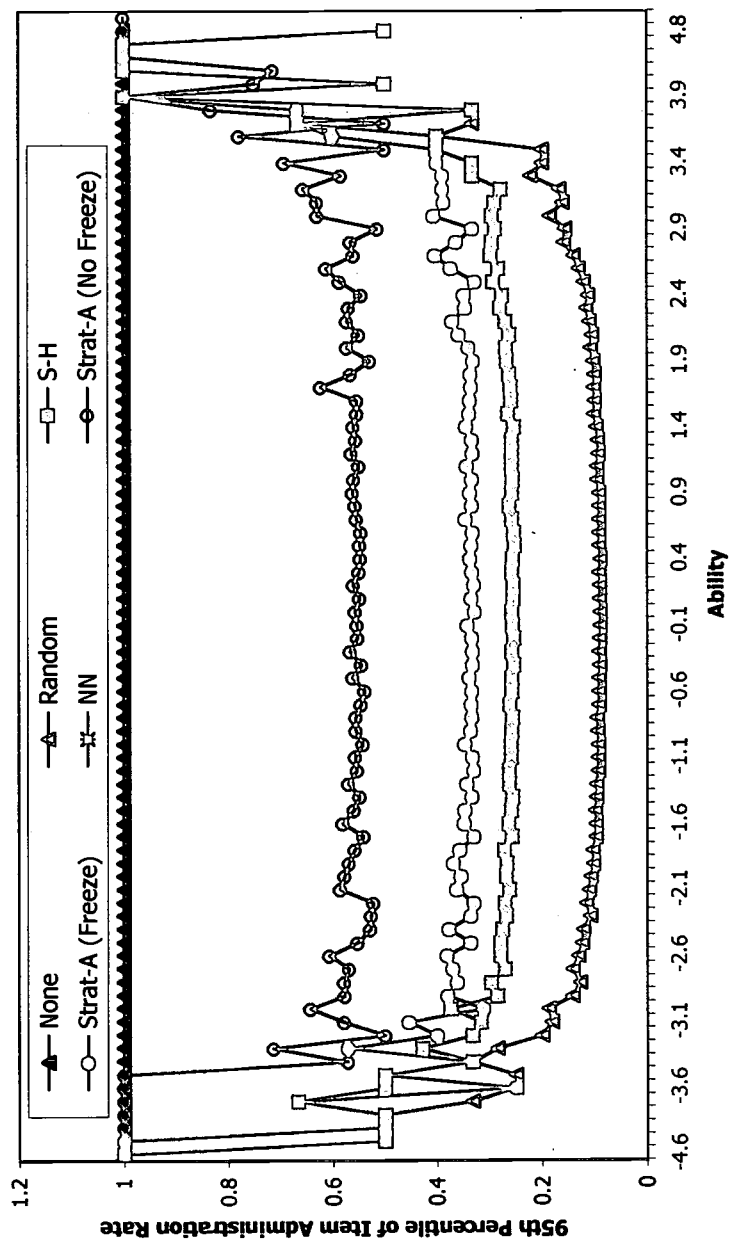
95th Percentiles of Conditional Exposure Distributions
Target Maximum Exposure Rate = 0.40

Figure 3 d

35

36
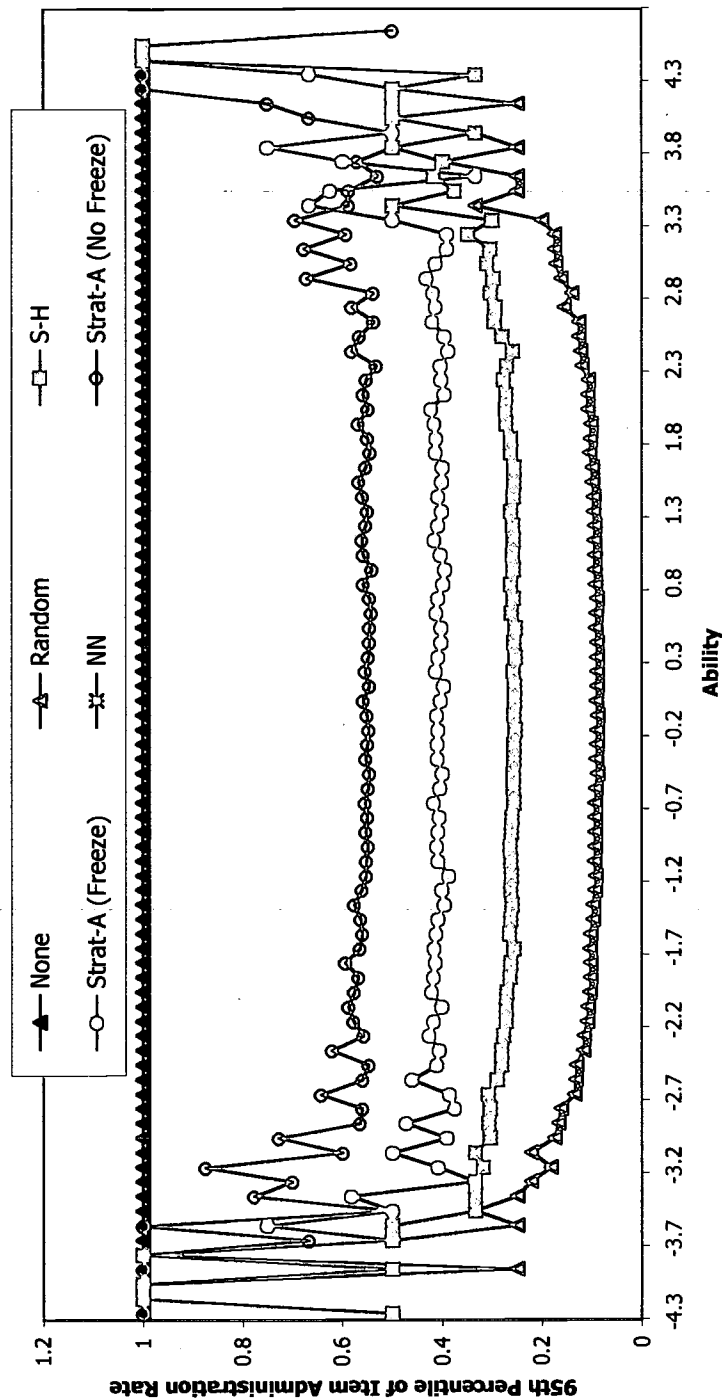
Posterior Variance of Ability Estimates: Target Freeze Rate = .15
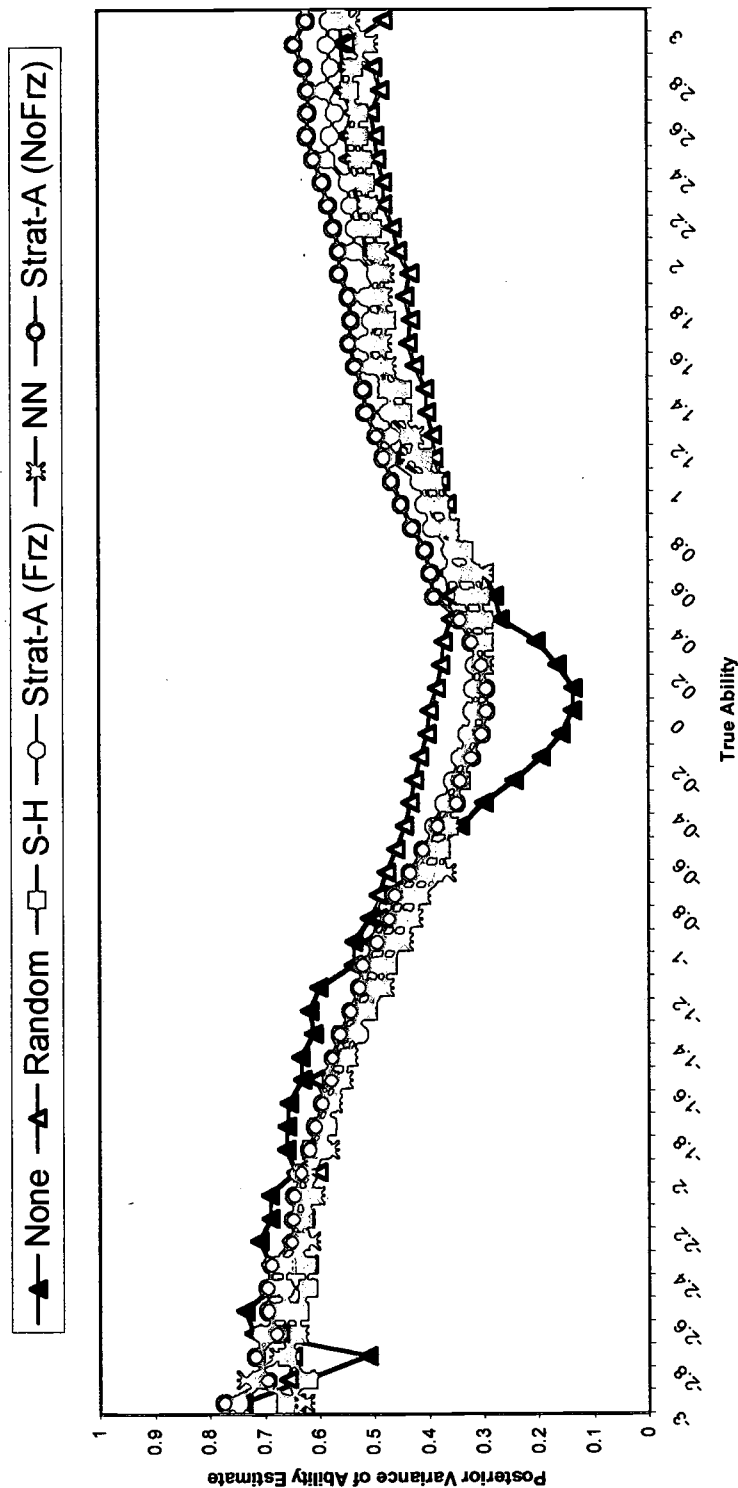
Figure 4 a

Posterior Variance of Ability Estimates: Target Freeze Rate = .25

Figure 4 b

39   40
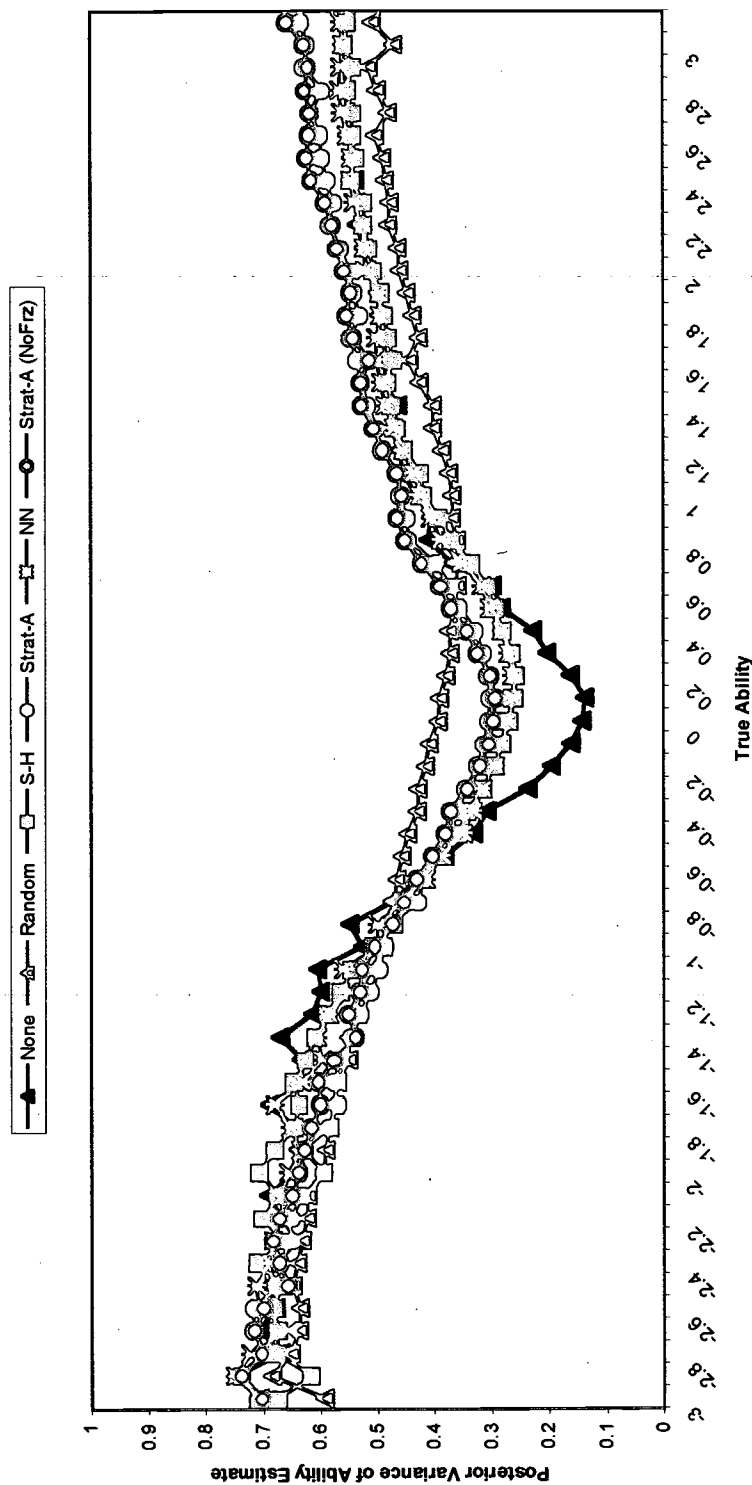
Posterior Variance of Ability Estimates: Target Freeze Rate = .33

Figure 4 c

41

42
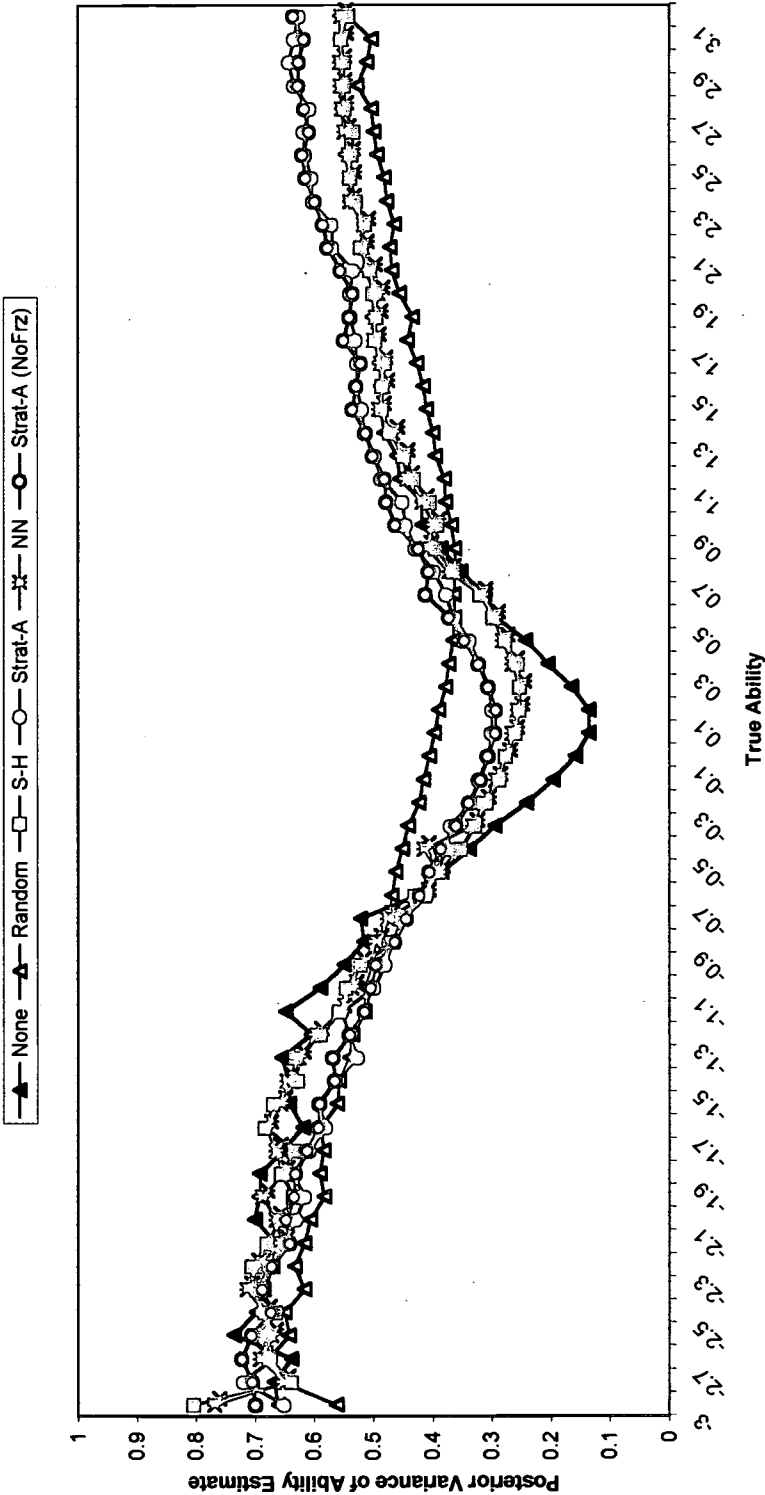
Posterior Variance of Ability Estimates: Target Freeze Rate = .40
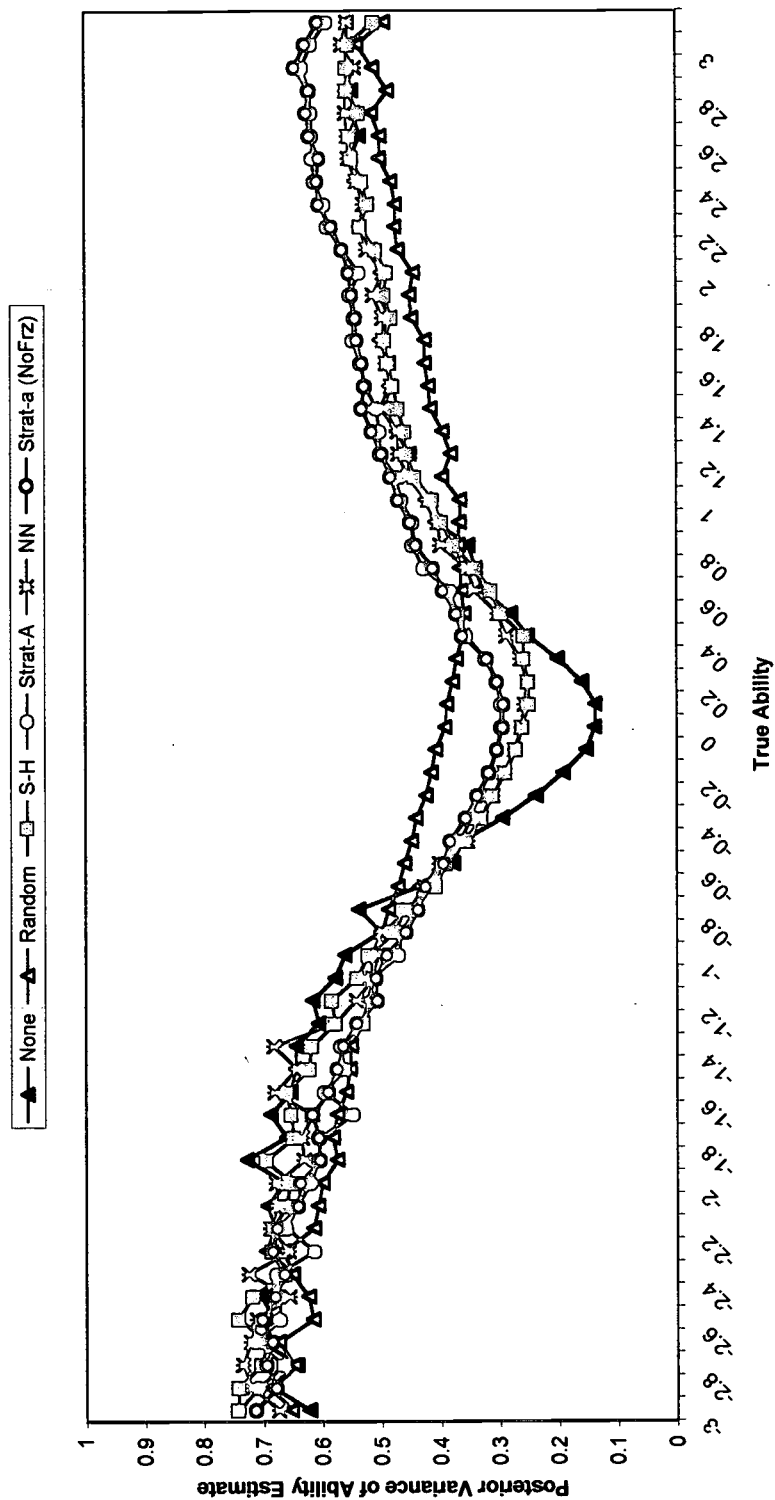
Figure 4 d
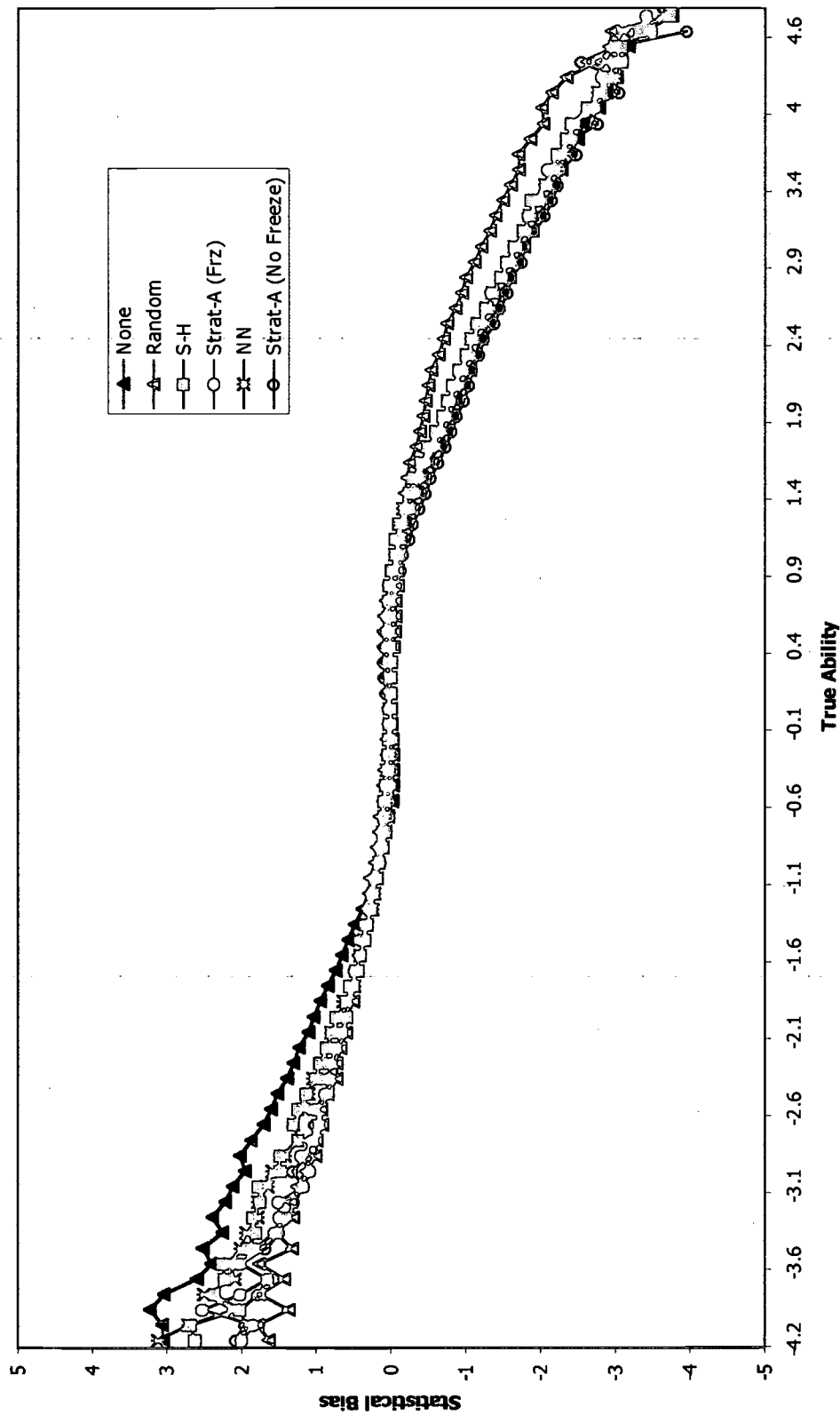
Figure 5 a

Bias in Ability Estimates Target Maximum Exposure Rate = 0.25

Figure 5 b

47

48
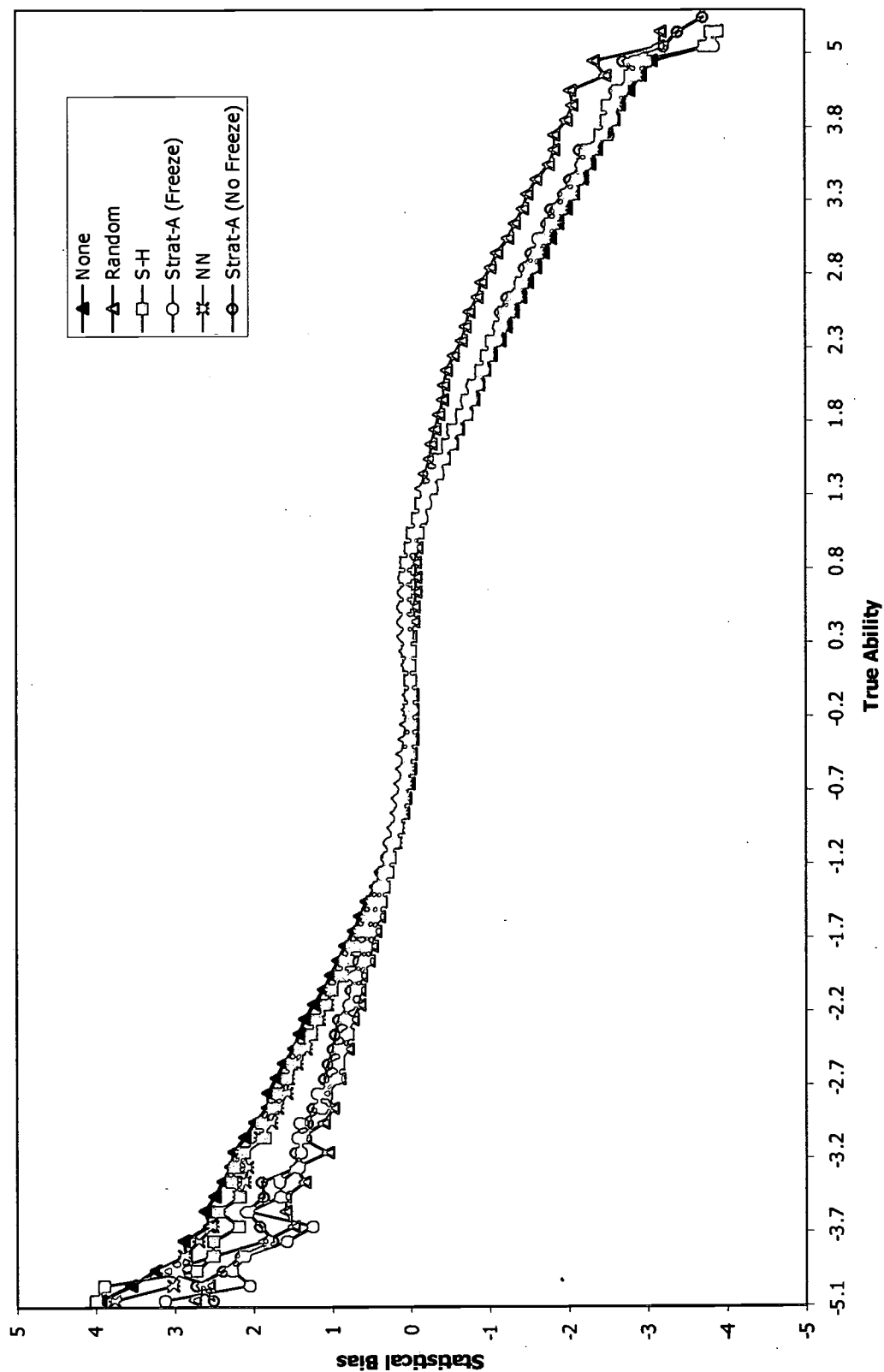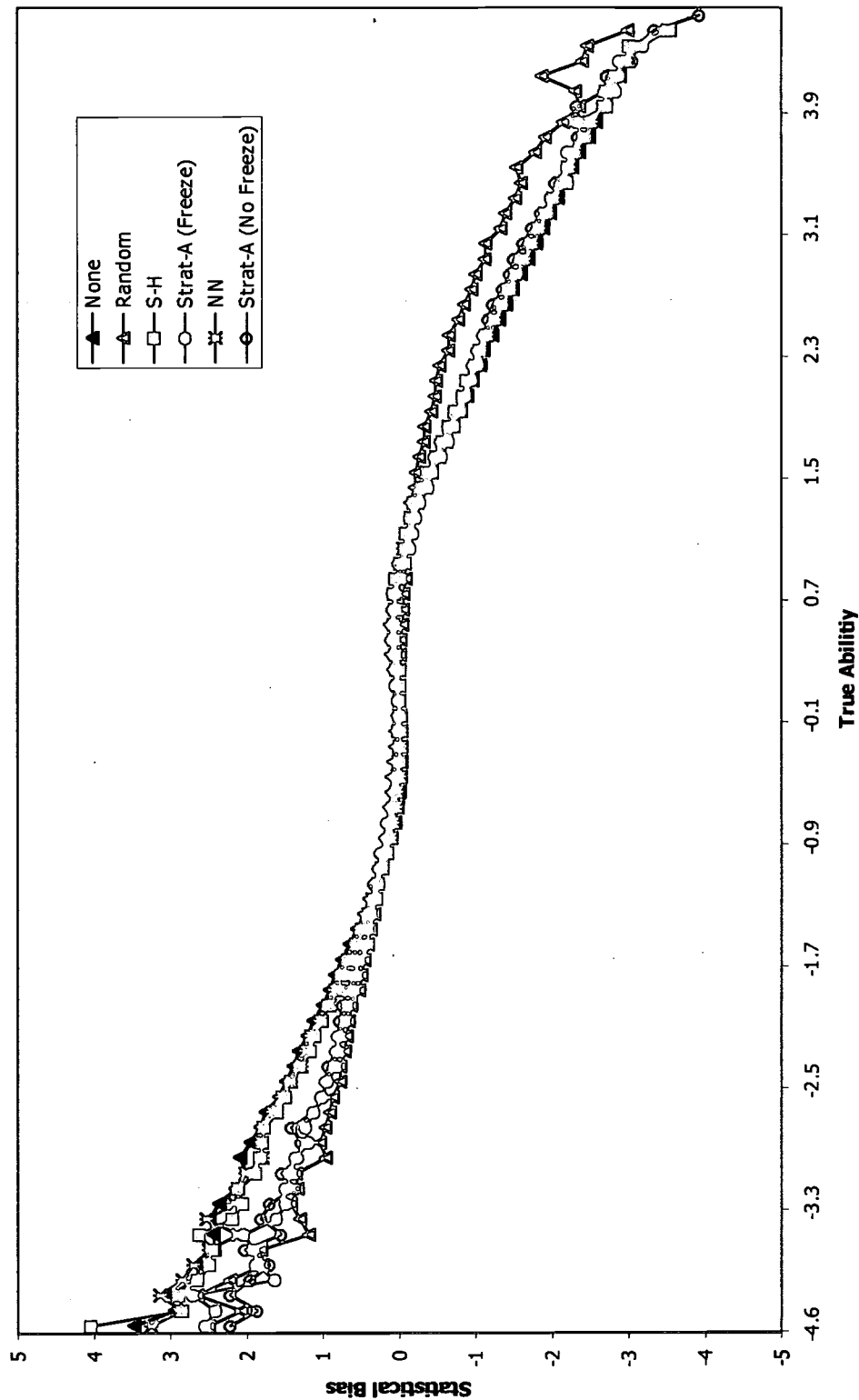
Figure 5 c

Bias in Ability Estimates Target Maximum Exposure Rate = 0.40
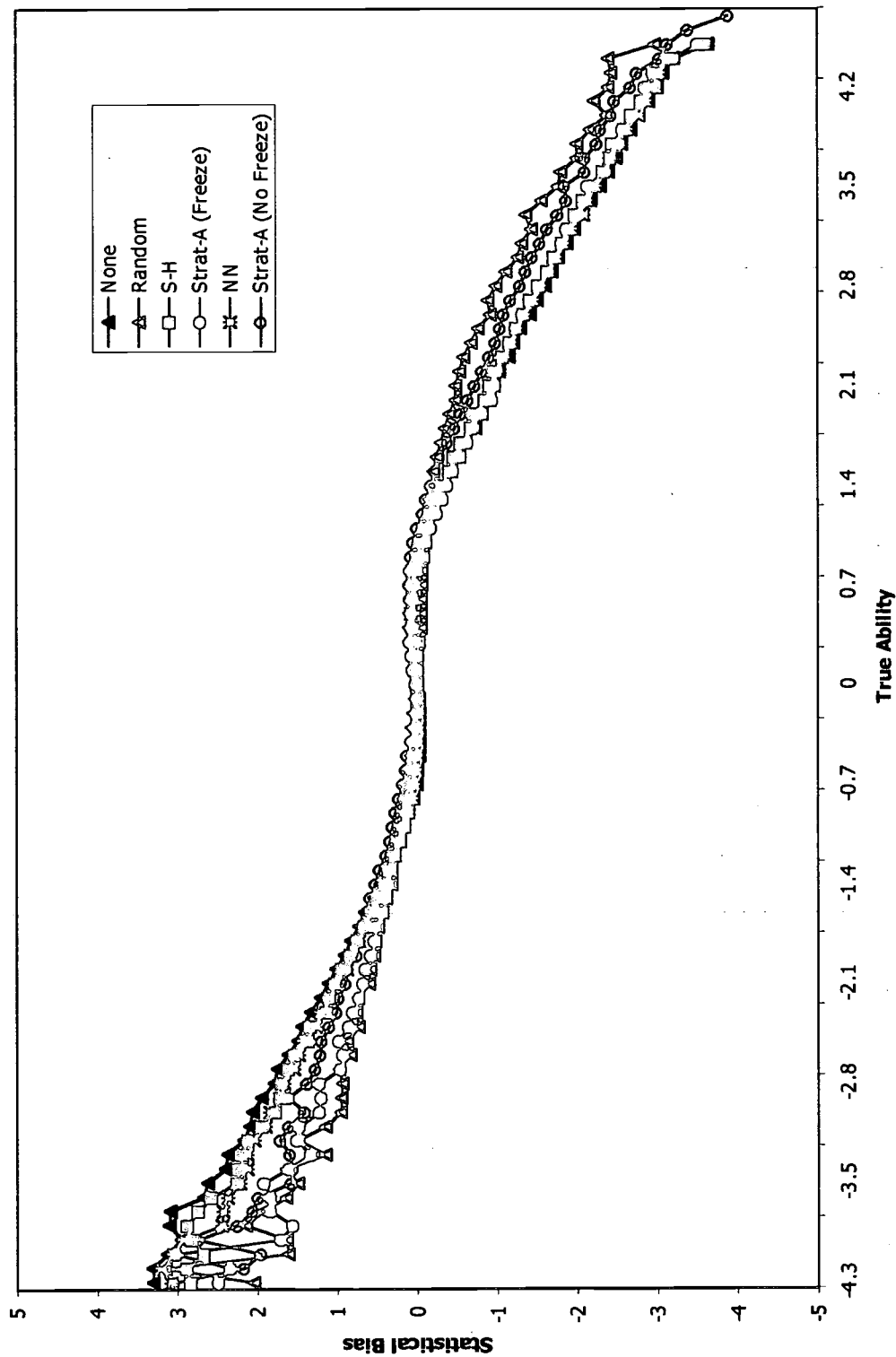
Figure 5 d
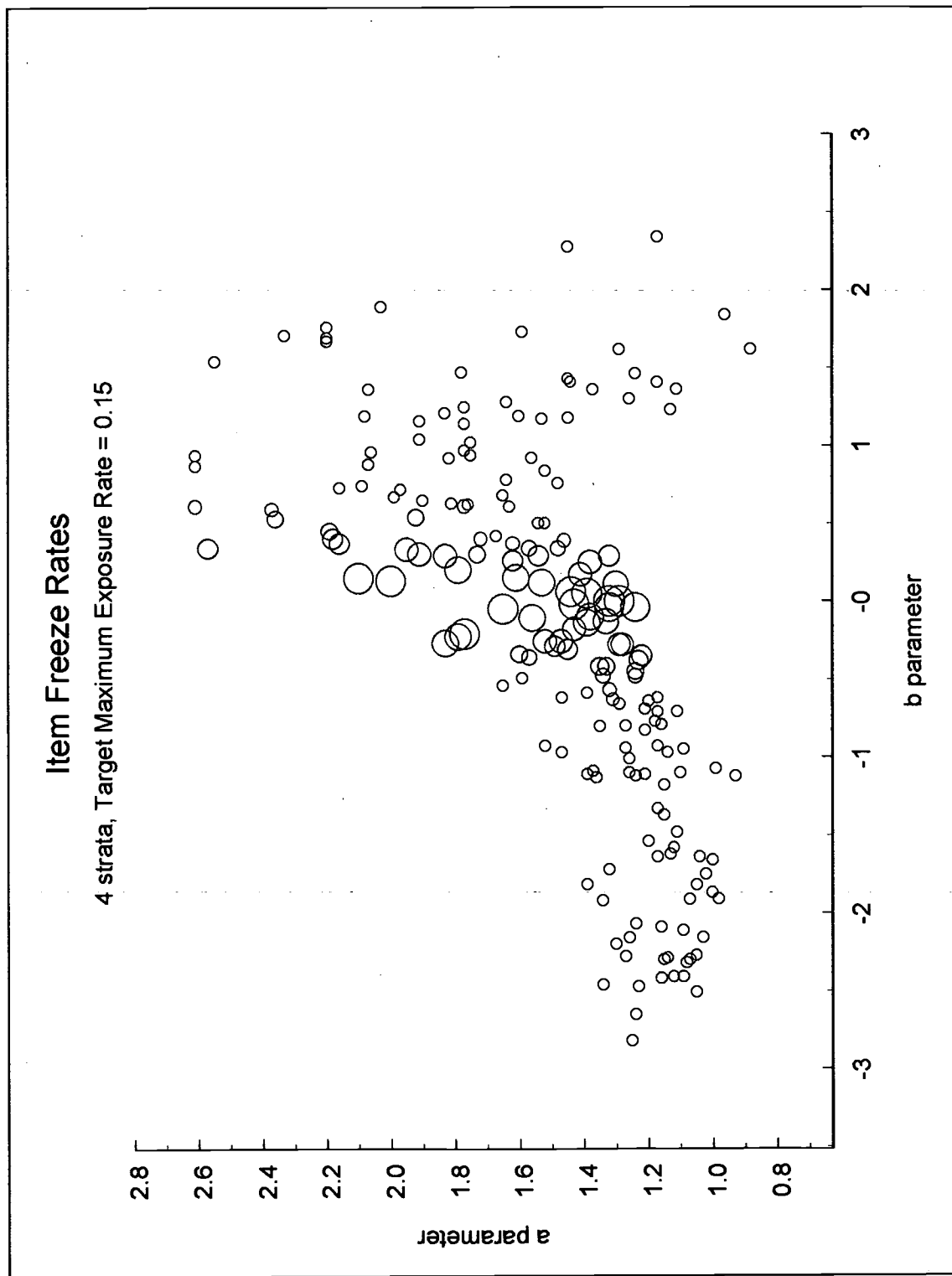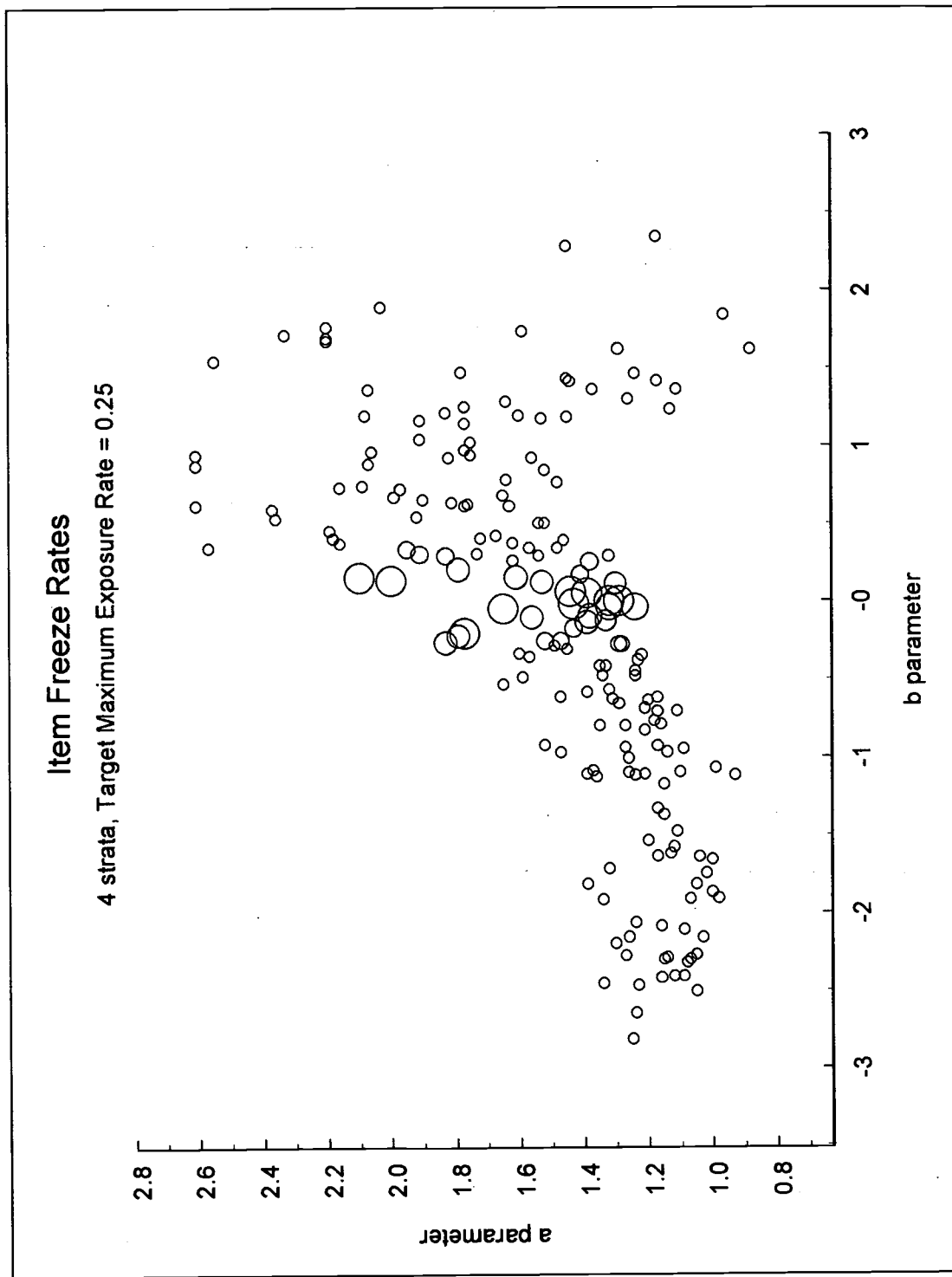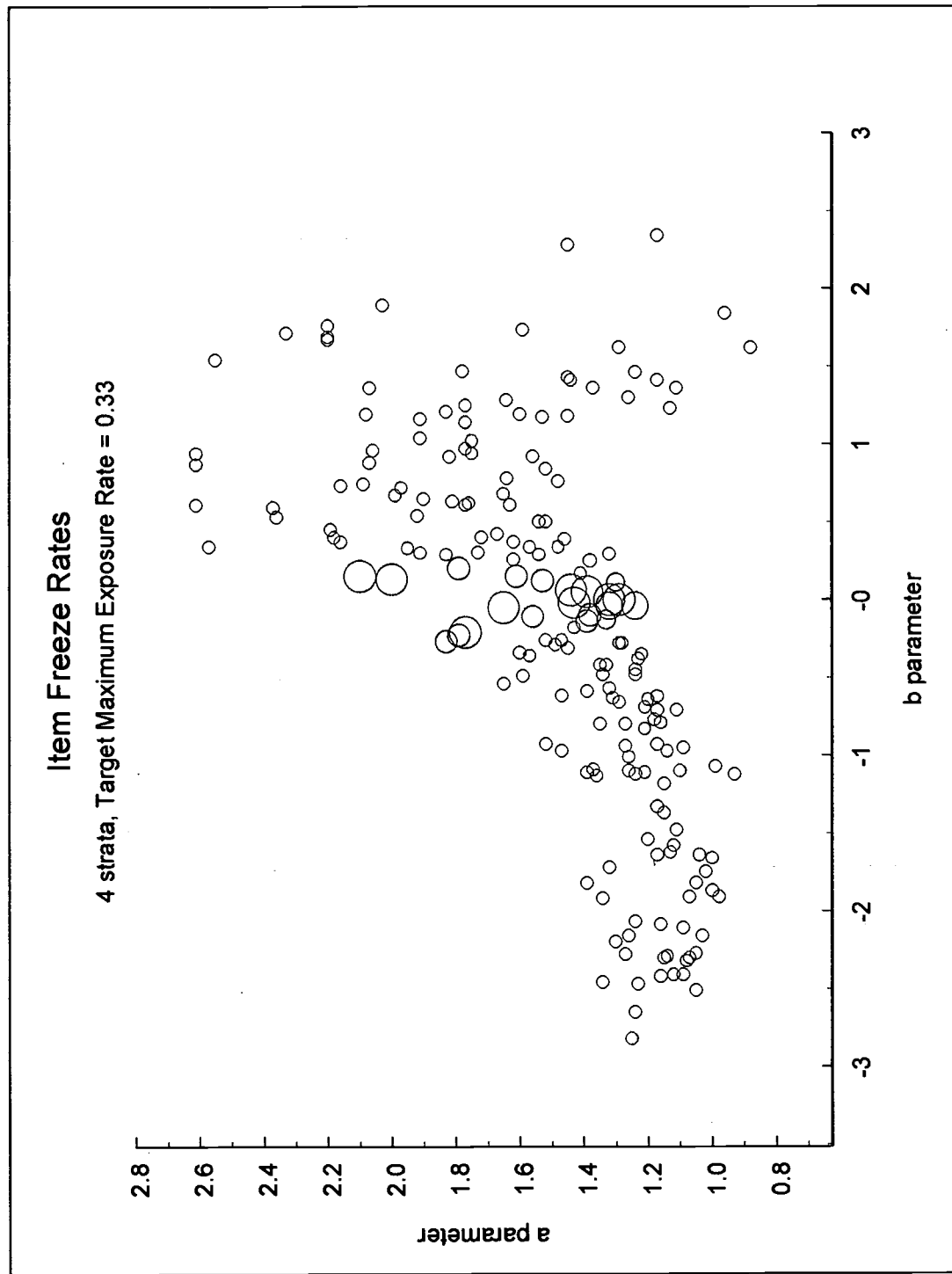
Item Freeze Rates

4 strata, Target Maximum Exposure Rate = 0.15

54

53

Figure 6 a

Item Freeze Rates

4 strata, Target Maximum Exposure Rate = 0.25

Figure 6 b

Figure 6 c
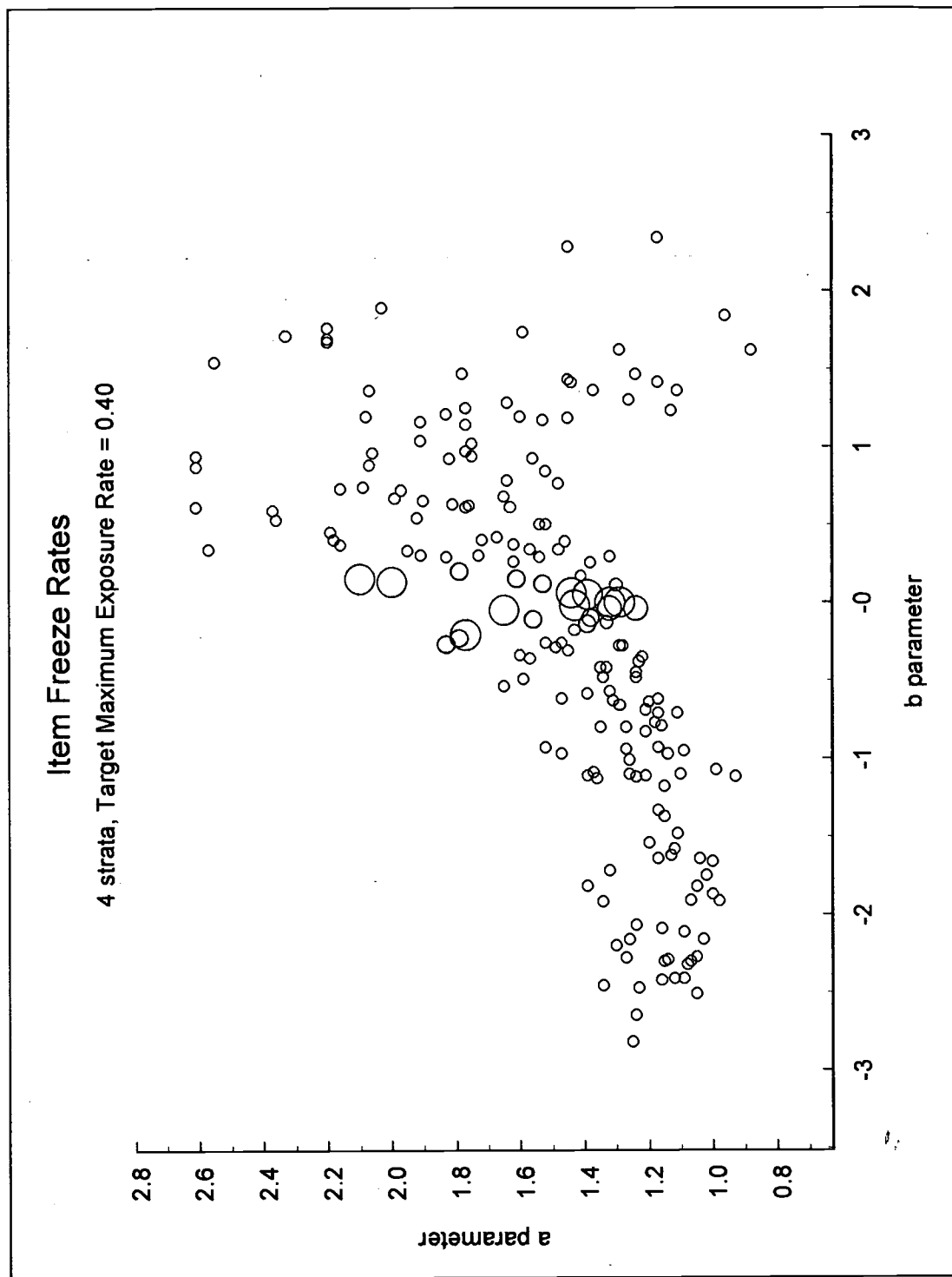
Item Freeze Rates

4 strata, Target Maximum Exposure Rate = 0.40

Figure 6 d

I. DOCUMENT IDENTIFICATION:

Title: *Nearest Neighbor, Simple Strata, + Probabilistic Parameters; An Empirical Comparison of Methods for Item Exposure Control in CATs*

Author(s): *Parshall, C.G, Kromrey, J. D., Harmes, J. C, + Sentovich, C.*

Corporate Source:

Publication Date: *NCME 2001*

II. REPRODUCTION RELEASE:

The sample sticker shown below will be affixed to all Level 1 documents

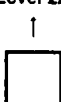PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

[✓]

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

[ ]

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

[ ]

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Sign here,→ please

Signature: *C. G. Parshall*

Organization/Address: *EDU 162, USF, Tampa, FL 33620*

Printed Name/Position/Title: *Dr. Cynthia G. Parshall*

Telephone: *813-974-1256* FAX: *813-974-4495*

E-Mail Address: *parshall@tempest.coedu.usf* Date: *5/15-01*

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

**University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions**

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com**

EFF-088 (Rev. 9/97)